

AK056137 Homo sapi  
BD176860 A method  
AL359588 Homo sapi  
AX877611 Sequence  
BD156750 Primer fo  
AK001821 Homo sapi  
BC006388 Mus muscu  
AC063950 Homo sapi  
AC005409 Homo sapi  
BC071687 Homo sapi  
CQ052066 Sequence  
CQ057110 Sequence  
CQ094163 Sequence  
CQ200647 Sequence  
CQ231806 Sequence  
CQ328863 Sequence  
CQ057085 Sequence  
CQ076348 Sequence  
CQ107339 Sequence  
CQ205801 Sequence  
CQ304361 Sequence  
CQ341629 Sequence  
L36434 Mus Musculu  
AX806612 Sequence  
AX869219 Sequence  
BD149281 Primer fo

20 2587.4 46.2 2603 9 AK056137  
21 2372.2 42.4 2428 6 BD176860  
22 2372.2 42.4 2428 9 HSM802690  
23 2018.2 36.1 2076 6 AX877611  
24 2018.2 36.1 2076 6 BD156750  
25 2018.2 36.1 2076 9 AK001821  
26 1795.4 32.1 4699 10 BC006388  
C 27 1716.4 30.7 174231 9 AC063950  
C 28 1716.4 30.7 177364 9 AC005409  
C 29 1544.2 27.6 1718 9 BC071687  
C 30 1104.8 19.7 1962 6 CQ052066  
C 31 1104.8 19.7 1962 6 CQ057110  
C 32 1104.8 19.7 1962 6 CQ094163  
C 33 1104.8 19.7 1962 6 CQ200647  
C 34 1104.8 19.7 1962 6 CQ231806  
C 35 1104.8 19.7 1962 6 CQ328863  
C 36 1103.4 19.7 1125 6 CQ057085  
C 37 1103.4 19.7 1125 6 CQ076348  
C 38 1103.4 19.7 1125 6 CQ107339  
C 39 1103.4 19.7 1125 6 CQ205801  
C 40 1103.4 19.7 1125 6 CQ304361  
C 41 1103.4 19.7 1125 6 CQ341629  
C 42 869 15.5 1846 10 MUSBDL2  
43 724.8 12.9 728 6 AX806612  
44 692 12.4 780 6 AX869219  
45 692 12.4 780 6 BD149281

ALIGNMENTS

RESULT 1  
AR300462 AR300462 5597 bp DNA linear PAT 12-JUN-2003  
LOCUS Sequence 4 from patent US 6537785.  
DEFINITION AR300462  
ACCESSION AR300462  
VERSION AR300462.1 GI:31687904  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 5597)  
AUTHORS Canfield, W.M.  
TITLE Methods of treating lysosomal storage diseases  
JOURNAL Patent: US 6537785-A 4 25-MAR-2003,  
FEATURES Location/Qualifiers  
source  
1..5597  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Query Match 100.0%; Score 5597; DB 6; Length 5597;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 CGGAGCCGAGCGGGCGTCCGTCCCGGAGCTGCAATGAGCGCGCCCGGAGGCTGTGACC 60  
Db 1 CGGAGCCGAGCGGGCGTCCGTCCCGGAGCTGCAATGAGCGCGCCCGGAGGCTGTGACC 60  
Qy 61 TGGCGGGCGGGCGCCGACCGGGCCCTGATGCGGGCTCGCTGAGCGCGGGCGGGCG 120  
Db 61 TGGCGGGCGGGCGCCGACCGGGCCCTGATGCGGGCTCGCTGAGCGCGGGCGGGCG 120  
Qy 121 GGGCGGCTCAGGCTCCCTCGGGCGCTGGCGGCTGAAGGGGTGATGCTGTTCAAGCTCC 180  
Db 121 GGGCGGCTCAGGCTCCCTCGGGCGCTGGCGGCTGAAGGGGTGATGCTGTTCAAGCTCC 180  
Qy 181 TGCAGAGACAAACCTATACCTGCTGTCCACAGGTATGGGCTCTACGTGTCTTTGG 240  
Db 181 TGCAGAGACAAACCTATACCTGCTGTCCACAGGTATGGGCTCTACGTGTCTTTGG 240  
Qy 241 GCCTCGTGTGACCATCGTCTCCGCTTCCAGTTCCGAGAGGTGTTCTTGGATGAGCC 300

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.  
OM nucleic - nucleic search, using sw model  
Run On: November 21, 2004, 20:12:11 ; Search time 22705 Seconds  
(without alignments)  
11657.364 Million cell updates/sec

Title: US-10-023-888-3  
Perfect score: 5597  
Sequence: 1 cggagccgagcggcgctccg.....aaaagttaattttgaaa 5597  
Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 4526729 seqs, 23644849745 residues  
Total number of hits satisfying chosen parameters: 9053458  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sv.\*  
13: gb\_un.\*  
14: gb\_vl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5597	100.0	5597	6 AR300462	AR300462 Sequence
2	5597	100.0	5597	6 AR428739	AR428739 Sequence
3	5597	100.0	5597	6 AR442828	AR442828 Sequence
4	4475.2	80.0	4511	9 AB033034	AB033034 Homo sapi
5	4326	77.3	4369	6 AX468102	AX468102 Sequence
6	4057	72.5	4277	6 AX468106	AX468106 Sequence
7	3761	67.2	3783	6 AR300471	AR300471 Sequence
8	3761	67.2	3783	6 AR428748	AR428748 Sequence
9	3761	67.2	3783	6 AR442837	AR442837 Sequence
10	3485.4	62.3	3621	6 AR300472	AR300472 Sequence
11	3485.4	62.3	3621	6 AR428749	AR428749 Sequence
12	3485.4	62.3	3621	6 AR442838	AR442838 Sequence
13	3100.2	55.4	3150	9 BC042615	BC042615 Homo sapi
14	2874.6	51.4	5204	10 AK171312	AK171312 Mus muscu
15	2820.8	50.4	5229	6 AR300468	AR300468 Sequence
16	2820.8	50.4	5229	6 AR428745	AR428745 Sequence
17	2820.8	50.4	5229	6 AR442834	AR442834 Sequence
18	2752.8	49.2	3152	6 CQ119596	CQ119596 Sequence
19	2587.4	46.2	2603	6 AX171469	AX171469 Sequence

Db 241 GCGTCGTTGTCACCATCGTCTCCGCTTCCAGTTCGGAGAGGTGGTTCTGGAATGAGCC 300  
Qy 301 GAGATCAATACCATGTTTTGTTGATTCCTATAGACAAATATTGCTGGAAAGTCCCTTC 360  
Db 301 GAGATCAATACCATGTTTTGTTGATTCCTATAGACAAATATTGCTGGAAAGTCCCTTC 360  
Qy 361 AGAATCGGCTTGTCTGCCATGCGGATGACGTTGTTTACACCTGGGTGAATGGCACAG 420  
Db 361 AGAATCGGCTTGTCTGCCATGCGGATGACGTTGTTTACACCTGGGTGAATGGCACAG 420  
Qy 421 ATCTTGAACCTACTGAAGAACTACAGCAGGTGACAGACAGATGGAGGAGGAGCAAG 480  
Db 421 ATCTTGAACCTACTGAAGAACTACAGCAGGTGACAGACAGATGGAGGAGGAGCAAG 480  
Qy 481 CAATGAGAGAAATCCTTGGGAAACACACGGAACCTTACTAAGAGAGTGAGAACGAGT 540  
Db 481 CAATGAGAGAAATCCTTGGGAAACACACGGAACCTTACTAAGAGAGTGAGAACGAGT 540  
Qy 541 TAGAGTGTTCGTAACACTGCAATTAAGGTGCCAATGCTTGTAAGTCCGACCCCTGC 600  
Db 541 TAGAGTGTTCGTAACACTGCAATTAAGGTGCCAATGCTTGTAAGTCCGACCCCTGC 600  
Qy 601 CAGCCAAACATCACCTGAGGAGCGTCCATCTCTTTTATCTCTTTTCAATCTGCCAGTG 660  
Db 601 CAGCCAAACATCACCTGAGGAGCGTCCATCTCTTTTATCTCTTTTCAATCTGCCAGTG 660  
Qy 661 ACATTTTCAATGTTGCAAAAACCAAAAACCCCTTCTACCAATGCTCAGTTGTTGTTTG 720  
Db 661 ACATTTTCAATGTTGCAAAAACCAAAAACCCCTTCTACCAATGCTCAGTTGTTGTTTG 720  
Qy 721 ACAGTACTAAGAGTGTGAAGATGCCACTCTGACACTGCTTAAAGGAAATAGCAGACAGA 780  
Db 721 ACAGTACTAAGAGTGTGAAGATGCCACTCTGACACTGCTTAAAGGAAATAGCAGACAGA 780  
Qy 781 CAGTATGAGGGGTACTTGACACAGATAAAGAGTCCCTGGATTAGTCTAATGCAAG 840  
Db 781 CAGTATGAGGGGTACTTGACACAGATAAAGAGTCCCTGGATTAGTCTAATGCAAG 840  
Qy 841 ATTTGGCTTCTGAGTGATTTCCACCAATTCAGGAAACAACTAATAACAA 900  
Db 841 ATTTGGCTTCTGAGTGATTTCCACCAATTCAGGAAACAACTAATAACAA 900  
Qy 901 AATTGCCAGAAAATCTTTCCTCTAAAGTCAAACTGTTGAGTGTGATTCAGAGGCCAGTG 960  
Db 901 AATTGCCAGAAAATCTTTCCTCTAAAGTCAAACTGTTGAGTGTGATTCAGAGGCCAGTG 960  
Qy 961 TAGCGCTTCTAAACTGAATTAACCCCAAGGATTTTCAAGAAATGAATTAAGCAAACTAAGA 1020  
Db 961 TAGCGCTTCTAAACTGAATTAACCCCAAGGATTTTCAAGAAATGAATTAAGCAAACTAAGA 1020  
Qy 1021 AGAATCAGCATGATGAGGAAAGAACTGACCAATAGTCTGATATTTATGCGATC 1080  
Db 1021 AGAATCAGCATGATGAGGAAAGAACTGACCAATAGTCTGATATTTATGCGATC 1080  
Qy 1081 TGAGGCCATCAGCGAGTCTAAGCAGGATGAAGACATCTCTGCCAGTCTGTTTGAAGATA 1140  
Db 1081 TGAGGCCATCAGCGAGTCTAAGCAGGATGAAGACATCTCTGCCAGTCTGTTTGAAGATA 1140  
Qy 1141 ACGAAGAACTGAGGTACTTATGCGATCTATCGAGAGGATGACCAATGCGGTTCCGGAATA 1200  
Db 1141 ACGAAGAACTGAGGTACTTATGCGATCTATCGAGAGGATGACCAATGCGGTTCCGGAATA 1200  
Qy 1201 TTTTCAATGTCACCAACGGGAGATTCATCTCGGCTGAACCTTGACAATCCTCGAGTGA 1260  
Db 1201 TTTTCAATGTCACCAACGGGAGATTCATCTCGGCTGAACCTTGACAATCCTCGAGTGA 1260  
Qy 1261 CAATAGTAACACACGAGATGTTTTCGAAATTTGAGCACTTGCCTACCTTTAGTTTAC 1320  
Db 1261 CAATAGTAACACACGAGATGTTTTCGAAATTTGAGCACTTGCCTACCTTTAGTTTAC 1320  
Qy 1321 CTGCTATTGAAAGTCAATTCATCGCATCGAAGGGCTGTCCCGAAGTCTTATTTACCTAA 1380  
Db 1321 CTGCTATTGAAAGTCAATTCATTCGCAATCGAAGGGCTGTCCCGAAGTCTTATTTACCTAA 1380

Qy 1381 ATGATGATGTCATGTTTGGAAAGGATGTCGCGCAGATGATTTTTTACAGTCACCTCAAAG 1440  
Db 1381 ATGATGATGTCATGTTTGGAAAGGATGTCGCGCAGATGATTTTTTACAGTCACCTCAAAG 1440  
Qy 1441 GCCAGAAGGTTTATTTGACATGCGCTGTGCCAAACTGTGCCAGGCTGCCAGGCTTCT 1500  
Db 1441 GCCAGAAGGTTTATTTGACATGCGCTGTGCCAAACTGTGCCAGGCTGCCAGGCTTCT 1500  
Qy 1501 GGATTAAGGATCGCTATTGTCAGAGGCTTGTAAATTTAGCCCTGCGATTTGGGATGGTG 1560  
Db 1501 GGATTAAGGATCGCTATTGTCAGAGGCTTGTAAATTTAGCCCTGCGATTTGGGATGGTG 1560  
Qy 1561 GGGATTGCTCTCGAAAACAGTGGAGGAGTCTGCTATTTATTCAGGAGGTGGAGTACTGGGA 1620  
Db 1561 GGGATTGCTCTCGAAAACAGTGGAGGAGTCTGCTATTTATTCAGGAGGTGGAGTACTGGGA 1620  
Qy 1621 GTATTGGAGTTGGACACCCCTGCGAGTTTGGTGGAGGAAATAAACAAGTCTCTTACTGTA 1680  
Db 1621 GTATTGGAGTTGGACACCCCTGCGAGTTTGGTGGAGGAAATAAACAAGTCTCTTACTGTA 1680  
Qy 1681 ATCAGGGATGTCGGAATTCCTGGCTCGCTGATTAAGTTCTGTGACCAAGCATGCAATGCT 1740  
Db 1681 ATCAGGGATGTCGGAATTCCTGGCTCGCTGATTAAGTTCTGTGACCAAGCATGCAATGCT 1740  
Qy 1741 TGTCTGTGGGTTGATGCTGGCGACTGTGGCGAGATCAATTTTCAATTTGTATAAAG 1800  
Db 1741 TGTCTGTGGGTTGATGCTGGCGACTGTGGCGAGATCAATTTTCAATTTGTATAAAG 1800  
Qy 1801 TGATCTCTTCCCAAAACAGACTCACTATATTTATTTCCAAAAGGTGAATGCTGCTTATT 1860  
Db 1801 TGATCTCTTCCCAAAACAGACTCACTATATTTATTTCCAAAAGGTGAATGCTGCTTATT 1860  
Qy 1861 TCAGCTTTGACAGAGTAGCAGGAGGTGAAGGTGCTTATAGTGACAAATCCAAATAA 1920  
Db 1861 TCAGCTTTGACAGAGTAGCAGGAGGTGAAGGTGCTTATAGTGACAAATCCAAATAA 1920  
Qy 1921 TTTGACATGCTTCTATTTGCAACAGTGAAGAAACCATCCACTCATATATGACAGTGAA 1980  
Db 1921 TTTGACATGCTTCTATTTGCAACAGTGAAGAAACCATCCACTCATATATGACAGTGAA 1980  
Qy 1981 TGAATGCCACCAACAATATTTTAACTCAAGTTTCAAAAATACAAAACGATGAAGAGTTCA 2040  
Db 1981 TGAATGCCACCAACAATATTTTAACTCAAGTTTCAAAAATACAAAACGATGAAGAGTTCA 2040  
Qy 2041 AAATCCAGATTAACAGTGGAGGTGACACAGGAGGACCAAACTGAATTTCTAGGCCC 2100  
Db 2041 AAATCCAGATTAACAGTGGAGGTGACACAGGAGGACCAAACTGAATTTCTAGGCCC 2100  
Qy 2101 AGAAGGGTTACGAAAATTTAGTTAGTCTCCATAAACACTTTTCCAGAGGCGGAAATCTTT 2160  
Db 2101 AGAAGGGTTACGAAAATTTAGTTAGTCTCCATAAACACTTTTCCAGAGGCGGAAATCTTT 2160  
Qy 2161 TTGAGGATATTTCCCAAGAAAACGTTTCCGAAAGTTTAAAGAGACATGATTTAACTCAA 2220  
Db 2161 TTGAGGATATTTCCCAAGAAAACGTTTCCGAAAGTTTAAAGAGACATGATTTAACTCAA 2220  
Qy 2221 CAAGGAGAGCCAGGAGGAGGTGAATTTCCCTGTTAAATTTTCACTCTCTTCCAAAAG 2280  
Db 2221 CAAGGAGAGCCAGGAGGAGGTGAATTTCCCTGTTAAATTTTCACTCTCTTCCAAAAG 2280  
Qy 2281 AGCCCGAGTTGAGTCTCAATACCTTTGGATTTGCAACTGGAACATGAGACATCACTTTGA 2340  
Db 2281 AGCCCGAGTTGAGTCTCAATACCTTTGGATTTGCAACTGGAACATGAGACATCACTTTGA 2340  
Qy 2341 AAGGATACAATTTGTCGAAGTCAAGCTTGGTGAATCATTTCTGATGAACATCAAGCATG 2400  
Db 2341 AAGGATACAATTTGTCGAAGTCAAGCTTGGTGAATCATTTCTGATGAACATCAAGCATG 2400  
Qy 2401 CTAAATTAATAAATAAAGTCTAATAACAGATGAATAAATAAAGTCTGCTGCTCCAC 2460  
Db 2401 CTAAATTAATAAATAAAGTCTAATAACAGATGAATAAATAAAGTCTGCTGCTCCAC 2460



2461 AGGAAAAACAGGTTTCATTAAGCATCTTGGCAACAGCTTAGGAGTGTCTGAAGATGTC 2520  
2461 AGGAAAAACAGGTTTCATTAAGCATCTTGGCAACAGCTTAGGAGTGTCTGAAGATGTC 2520  
2521 AGAGGTTCACATTTTCTCCAGTGTAGTGTAAAGTGAATGGTCAATGACAGGCTCAGAAATC 2580  
2521 AGAGGTTCACATTTTCTCCAGTGTAGTGTAAAGTGAATGGTCAATGACAGGCTCAGAAATC 2580  
2581 CACCCCTGGACTTGGAGACCAAGCAAGATTTAGAGTGGAAATCTCACACCCCAAAAAACCA 2640  
2581 CACCCCTGGACTTGGAGACCAAGCAAGATTTAGAGTGGAAATCTCACACCCCAAAAAACCA 2640  
2641 TAGGGGAAATGTGAAGAAAGAAAGCCCAATCTCTGATTTGCTTCCACTGGAAGCCAGA 2700  
2641 TAGGGGAAATGTGAAGAAAGAAAGCCCAATCTCTGATTTGCTTCCACTGGAAGCCAGA 2700  
2701 TGACAAAAAGAAAGAAATTCACAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAATG 2760  
2701 TGACAAAAAGAAAGAAATTCACAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAATG 2760  
2761 CTGAAGATTCATAGGCGTTTACTGAGTGTACTTGGAGAAAGAGTGCAGCATTTACACAG 2820  
2761 CTGAAGATTCATAGGCGTTTACTGAGTGTACTTGGAGAAAGAGTGCAGCATTTACACAG 2820  
2821 ATAGTTACTTTGGGCTTTTGGCATGGGAGAAAGAAAGTATTTCCAGATCTTCTCGAGC 2880  
2821 ATAGTTACTTTGGGCTTTTGGCATGGGAGAAAGAAAGTATTTCCAGATCTTCTCGAGC 2880  
2881 AAGAGAGTCATTAAGACACAAATTTGGCATCTTCACTGTAGCAAAAAATCTGGGAGGC 2940  
2881 AAGAGAGTCATTAAGACACAAATTTGGCATCTTCACTGTAGCAAAAAATCTGGGAGGC 2940  
2941 AACTAAAGATACATTTCCAGATTTCCCTCAGATATGTAATAAATTTCTAAATAGCAAGT 3000  
2941 AACTAAAGATACATTTCCAGATTTCCCTCAGATATGTAATAAATTTCTAAATAGCAAGT 3000  
3001 TTGGATTACATCGCGGAAAGTCCCTGCTCAGATGCTCAGATGTAACCGGATTTGTTA 3060  
3001 TTGGATTACATCGCGGAAAGTCCCTGCTCAGATGCTCAGATGTAACCGGATTTGTTA 3060  
3061 TGCAGAACTGCAAGATATGTTCCCTGAAAGATTTTGAAGACAGTCAATTCACAAAGTGC 3120  
3061 TGCAGAACTGCAAGATATGTTCCCTGAAAGATTTTGAAGACAGTCAATTCACAAAGTGC 3120  
3121 GCATTTCTGAGATATGAGTTTGGCTTCTCTTATTTTATTTATCTATGAGTGCAGTGC 3180  
3121 GCATTTCTGAGATATGAGTTTGGCTTCTCTTATTTTATTTATTTATCTATGAGTGCAGTGC 3180  
3181 AGCCACTGAATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTGTCTTGT 3240  
3181 AGCCACTGAATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTGTCTTGT 3240  
3241 CTGACAGAAATCCGAACACTGGCTACAGAAATTCAGAACTGCGGTTAAGTTTGCAGG 3300  
3241 CTGACAGAAATCCGAACACTGGCTACAGAAATTCAGAACTGCGGTTAAGTTTGCAGG 3300  
3301 ATTGACAGGTCTGGAACACATGCTAATAAATTTGCTCAAAAAATGCTTCTGCTGATATCA 3360  
3301 ATTGACAGGTCTGGAACACATGCTAATAAATTTGCTCAAAAAATGCTTCTGCTGATATCA 3360  
3361 CGCAGCTAATAATATTCACCACTCCAGATCTCTATGATCCCACTGCCACCGG 3420  
3361 CGCAGCTAATAATATTCACCACTCCAGATCTCTATGATCCCACTGCCACCGG 3420  
3421 TCACTAAAAGTCTAGTAAACAACTGTAAACCACTGAATCTGACAAATTCACAAAGCATATA 3480  
3421 TCACTAAAAGTCTAGTAAACAACTGTAAACCACTGAATCTGACAAATTCACAAAGCATATA 3480  
3481 AGGACAAAAACAAATATAGTTTGAATCATGGGAGAAAGAAATCGGTTTTAAATGA 3540  
3481 AGGACAAAAACAAATATAGTTTGAATCATGGGAGAAAGAAATCGGTTTTAAATGA 3540  
3541 TTCGTACCAACGTTTCTCATGTGGTGGCCAGTTGGATGACATAAGAAAAACCCCTAGGA 3600

3541 TTCTGTACCAACGTTTCTCATGTGGTGGCCAGTTGGATGACATAGAAAAAACCCCTAGGA 3600  
3601 AGTTTGTTCCTCGAATGAACAATTCAGCAACCAATCATAAAGATGCTCAGACAGTGAAGG 3660  
3601 AGTTTGTTCCTCGAATGAACAATTCAGCAACCAATCATAAAGATGCTCAGACAGTGAAGG 3660  
3661 CTGTTCTCAGGACCTTCTATGATGATCCATCTTCCCATACCTTCCCAATTTGAGTCCCAA 3720  
3661 CTGTTCTCAGGACCTTCTATGATGATCCATCTTCCCATACCTTCCCAATTTGAGTCCCAA 3720  
3721 GAGAGTATCGAAACCGGTTTCTTTCATATGATGAGCTGAGGAATGGAGGCTTATTCGAG 3780  
3721 GAGAGTATCGAAACCGGTTTCTTTCATATGATGAGCTGAGGAATGGAGGCTTATTCGAG 3780  
3781 ACAAATTAAGATTTTGGACCCATTTGTTACTAGCAACATGATTTTCTTACTATATCTT 3840  
3781 ACAAATTAAGATTTTGGACCCATTTGTTACTAGCAACATGATTTTCTTACTATATCTT 3840  
3841 CATTTTGTCTGAGCAGTTAAATTCACCTTAAGCGGAAGATATTTCCCAAGAGGAGATAC 3900  
3841 CATTTTGTCTGAGCAGTTAAATTCACCTTAAGCGGAAGATATTTCCCAAGAGGAGATAC 3900  
3901 ACAAAGAGCTAGTCCCAATTCGAATCAGAGTATAGAGATCTTTCATTTGAAAAACCATCTA 3960  
3901 ACAAAGAGCTAGTCCCAATTCGAATCAGAGTATAGAGATCTTTCATTTGAAAAACCATCTA 3960  
3961 CCTCAGCATTTACTGAGCATTTTAAACTCAGCTTCCAGAGATGCTTTTGTGATGTAT 4020  
3961 CCTCAGCATTTACTGAGCATTTTAAACTCAGCTTCCAGAGATGCTTTTGTGATGTAT 4020  
4021 GCTTAGCAGTTTGGCCCGGAGAAAGAAATATCCAGTACCATCTGTTTGTGCGCATGAA 4080  
4021 GCTTAGCAGTTTGGCCCGGAGAAAGAAATATCCAGTACCATCTGTTTGTGCGCATGAA 4080  
4081 TATAGCCACTGACTAGGAATTTATTTAAACCAACCACTGAAACTTGTGTGCGAGCAGC 4140  
4081 TATAGCCACTGACTAGGAATTTATTTAAACCAACCACTGAAACTTGTGTGCGAGCAGC 4140  
4141 TCTGAACTGATTTTACTTTTAAAGAAATTTGCTCATGAGCTGCTCATCTTTTATATAAAA 4200  
4141 TCTGAACTGATTTTACTTTTAAAGAAATTTGCTCATGAGCTGCTCATCTTTTATATAAAA 4200  
4201 GGCTCACTGACAAGAGACAGCTGTTAAATTTCCCAAGCAATCATTTGAGACTAACCTTAT 4260  
4201 GGCTCACTGACAAGAGACAGCTGTTAAATTTCCCAAGCAATCATTTGAGACTAACCTTAT 4260  
4261 TAGGAGAGCTATGCCAGCTGGAGTGTGCTTAAGAGGCTCCAGTCTTTGCAATTCCAA 4320  
4261 TAGGAGAGCTATGCCAGCTGGAGTGTGCTTAAGAGGCTCCAGTCTTTGCAATTCCAA 4320  
4321 AGCCTTTTGTAAAGTTTTCACCTTTTTCATTTTCCCAATTTTAAAGTAGTACTA 4380  
4321 AGCCTTTTGTAAAGTTTTCACCTTTTTCATTTTCCCAATTTTAAAGTAGTACTA 4380  
4381 AGTTAACTAGTATTCTTCTGCTTCTGATATAAGAAATTTGGATGCTTAAACCTATTTTA 4440  
4381 AGTTAACTAGTATTCTTCTGCTTCTGATATAAGAAATTTGGATGCTTAAACCTATTTTA 4440  
4441 TAGATGTTATTTAAATTAATGACCAATATCACCTCTTATTGACAATACCTAAATTAAG 4500  
4441 TAGATGTTATTTAAATTAATGACCAATATCACCTCTTATTGACAATACCTAAATTAAG 4500  
4501 TTTTATTTAAATTTAAGACTGTAAGTCTTAAACCACTTAACCTACTGAAAGCTCAATG 4560  
4501 TTTTATTTAAATTTAAGACTGTAAGTCTTAAACCACTTAACCTACTGAAAGCTCAATG 4560  
4561 ATTGACATCTGAATGCTTTGTAATTTGACTTTCAGCTTAAAGATGCTATGATTTCA 4620  
4561 ATTGACATCTGAATGCTTTGTAATTTGACTTTCAGCTTAAAGATGCTATGATTTCA 4620  
4621 CGTGCAGGCTTAATTTCAACAGGCTAGATGATGATCTTACAGATGTAATGATTTT 4680



Qy 841 ATTTGGCTTTCCTGAGTGGATTTCCACCAACATTCAGGAACAAATCAACTAAACAA 900  
Db 841 ATTTGGCTTTCCTGAGTGGATTTCCACCAACATTCAGGAACAAATCAACTAAACAA 900  
Qy 901 AATTCCGAGAAATCTTTCTCTAAAGTCAAACTGTTCAGTGTATTCAGAGCCAGTG 960  
Db 901 AATTCCGAGAAATCTTTCTCTAAAGTCAAACTGTTCAGTGTATTCAGAGCCAGTG 960  
Qy 961 TAGCCCTTCTAAACCTGAATACCCCAAGGATTTTCAGGAATTCAGTAAGCAACTAAGA 1020  
Db 961 TAGCCCTTCTAAACCTGAATACCCCAAGGATTTTCAGGAATTCAGTAAGCAACTAAGA 1020  
Qy 1021 AGAACATGACCATTCATGAGGAAAGAACTGACCAATAGTCTCGCATATTTATTCGGATC 1080  
Db 1021 AGAACATGACCATTCATGAGGAAAGAACTGACCAATAGTCTCGCATATTTATTCGGATC 1080  
Qy 1081 TGAGGCCATGAGCAGTCTAAGCAGGATGAAGACATCTCTGCGAGTGTTCAGAGATA 1140  
Db 1081 TGAGGCCATGAGCAGTCTAAGCAGGATGAAGACATCTCTGCGAGTGTTCAGAGATA 1140  
Qy 1141 ACGAAGAACTGAGGTACTCATTCGATCTATCGAGAGCATGACCATGGGTTCGGAATA 1200  
Db 1141 ACGAAGAACTGAGGTACTCATTCGATCTATCGAGAGCATGACCATGGGTTCGGAATA 1200  
Qy 1201 TTTTCATTGTACCAACGGGAGATTCATCTGCTGCTGAACCTTTGAACAATCTCGAGTGA 1260  
Db 1201 TTTTCATTGTACCAACGGGAGATTCATCTGCTGCTGAACCTTTGAACAATCTCGAGTGA 1260  
Qy 1261 CAATAGTAAACACACAGAGATTTTCGAAATTTGAGCCACTTGCTACCTTTAGTTCAC 1320  
Db 1261 CAATAGTAAACACACAGAGATTTTCGAAATTTGAGCCACTTGCTACCTTTAGTTCAC 1320  
Qy 1321 CTGCTATTGAAAGTCAATTCATCGCATCGAAGGCTGTCCAGAAAGTTATTTACCTAA 1380  
Db 1321 CTGCTATTGAAAGTCAATTCATCGCATCGAAGGCTGTCCAGAAAGTTATTTACCTAA 1380  
Qy 1381 ATGATGATGCTATTTGGGAGGATGTCTGCGCAGATGATTTTACAGTCACTCCAAAG 1440  
Db 1381 ATGATGATGCTATTTGGGAGGATGTCTGCGCAGATGATTTTACAGTCACTCCAAAG 1440  
Qy 1441 GCCAAGGTTTATTTGACATGGCTGTGCCAACTGTGCGAGGCTGCCAGGTTCTT 1500  
Db 1441 GCCAAGGTTTATTTGACATGGCTGTGCCAACTGTGCGAGGCTGCCAGGTTCTT 1500  
Qy 1501 GGATTAAGATGGCTATTGTGCAAGGCTGTGATTAATTCAGCTGCGATTCGGATGGTG 1560  
Db 1501 GGATTAAGATGGCTATTGTGCAAGGCTGTGATTAATTCAGCTGCGATTCGGATGGTG 1560  
Qy 1561 GGGATTGCTCTGGAACACAGTGGAGGAGTCCGTATTTGCGAGGCTGGAGTACTGGGA 1620  
Db 1561 GGGATTGCTCTGGAACACAGTGGAGGAGTCCGTATTTGCGAGGCTGGAGTACTGGGA 1620  
Qy 1621 GTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAACAGTGTCTCTTACTGTA 1680  
Db 1621 GTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAACAGTGTCTCTTACTGTA 1680  
Qy 1681 ATCAGGGATGTCGGAATTCCTGGCTCGCTGATTAAGTCTGTGACCAAGCATGCAATGCT 1740  
Db 1681 ATCAGGGATGTCGGAATTCCTGGCTCGCTGATTAAGTCTGTGACCAAGCATGCAATGCT 1740  
Qy 1741 TGTCTGTGGGTTTGTATGCTGGGAGTGTGGGCAAGATCAATTTTCATGAATTTGATAAG 1800  
Db 1741 TGTCTGTGGGTTTGTATGCTGGGAGTGTGGGCAAGATCAATTTTCATGAATTTGATAAG 1800  
Qy 1801 TGAATCTTCTCCAAACAGATCTCATATATTTATTCBAAAAGGTGAATGCCCTTATTT 1860  
Db 1801 TGAATCTTCTCCAAACAGATCTCATATATTTATTCBAAAAGGTGAATGCCCTTATTT 1860  
Qy 1861 TCAGCTTTTCGAGAGTAGCCAAAGAGGAGTTGAAGGTGCTTATAGTGCATTCACATAA 1920  
Db 1861 TCAGCTTTTCGAGAGTAGCCAAAGAGGAGTTGAAGGTGCTTATAGTGCATTCACATAA 1920

Qy 1921 TTGCAGATGCTTCTATTGGCAACAAGTGGAAACCATTCACCTCATAATTCACAGTGGAA 1980  
Db 1921 TTGCAGATGCTTCTATTGGCAACAAGTGGAAACCATTCACCTCATAATTCACAGTGGAA 1980  
Qy 1981 TGAATGCCACCAACATTAATTTAATCTCAGTTTCAAAATACAAACGATGAAGATTCA 2040  
Db 1981 TGAATGCCACCAACATTAATTTAATCTCAGTTTCAAAATACAAACGATGAAGATTCA 2040  
Qy 2041 AAATGCAGATAACAGTGGAGGTGGACACAAGGAGGAGCAAAACCTGAATTTCTACGGCCC 2100  
Db 2041 AAATGCAGATAACAGTGGAGGTGGACACAAGGAGGAGCAAAACCTGAATTTCTACGGCCC 2100  
Qy 2101 AGAAGGTTACGAAAATTTAGTTAGTTCCTCATAACTCTTTCAGAGGCGGAAATCCTTT 2160  
Db 2101 AGAAGGTTACGAAAATTTAGTTAGTTCCTCATAACTCTTTCAGAGGCGGAAATCCTTT 2160  
Qy 2161 TTGAGGATATTCCTCAAGAAACGCTTCCCGAAGTTTAAAGACATGATTTAACTCAA 2220  
Db 2161 TTGAGGATATTCCTCAAGAAACGCTTCCCGAAGTTTAAAGACATGATTTAACTCAA 2220  
Qy 2221 CAAGGAGAGCCAGGAGGAGTGAATAATTCCTGTGTAATTTCTCACTCCTTCCAAAAG 2280  
Db 2221 CAAGGAGAGCCAGGAGGAGTGAATAATTCCTGTGTAATTTCTCACTCCTTCCAAAAG 2280  
Qy 2281 ACSCCCAGTTGAGTCTCAATACCTTGGATTTGCACTGGAAACATGGAGACATCCTTTGA 2340  
Db 2281 ACSCCCAGTTGAGTCTCAATACCTTGGATTTGCACTGGAAACATGGAGACATCCTTTGA 2340  
Qy 2341 AAGGATACAAATTTGTCCAGTCAGCTTGTGAGATCAATTTCTGATGAACTCACAGCATG 2400  
Db 2341 AAGGATACAAATTTGTCCAGTCAGCTTGTGAGATCAATTTCTGATGAACTCACAGCATG 2400  
Qy 2401 CTAAATTAATAATCAAGCTATTAATAACAGATGAACAAATGACAGTGTGGTGCTCCAC 2460  
Db 2401 CTAAATTAATAATCAAGCTATTAATAACAGATGAACAAATGACAGTGTGGTGCTCCAC 2460  
Qy 2461 AGGAAACACAGTTTCAAAAAGCATCTTGCCAAACAGCTTAGGAGTGTCTGAAAGATTGC 2520  
Db 2461 AGGAAACACAGTTTCAAAAAGCATCTTGCCAAACAGCTTAGGAGTGTCTGAAAGATTGC 2520  
Qy 2521 AGAGGTTGACTTTTCTGCAAGTGAATTAAGTGAATGGTCAATGACAGGTTCAAAATC 2580  
Db 2521 AGAGGTTGACTTTTCTGCAAGTGAATTAAGTGAATGGTCAATGACAGGTTCAAAATC 2580  
Qy 2581 CACCCCTGGACTTTGGAGACACAGCAAGATTTAGAGTGAAACTCACACCAAAACCA 2640  
Db 2581 CACCCCTGGACTTTGGAGACACAGCAAGATTTAGAGTGAAACTCACACCAAAACCA 2640  
Qy 2641 TAGCGGAAATGTGACAAAGAAAGCCCTCATCTCTGATTTGTTCCACTGGAAAGCCAGA 2700  
Db 2641 TAGCGGAAATGTGACAAAGAAAGCCCTCATCTCTGATTTGTTCCACTGGAAAGCCAGA 2700  
Qy 2701 TGACAAAGAAAGAAATTCACAGGAAAGAAAGACAGTAAAGTGGAGAAATG 2760  
Db 2701 TGACAAAGAAAGAAATTCACAGGAAAGAAAGACAGTAAAGTGGAGAAATG 2760  
Qy 2761 CTGAAATACATAGGCTTACTGAAAGTGTACTTGGAAAGAAAGCTGCAGCATTAACAG 2820  
Db 2761 CTGAAATACATAGGCTTACTGAAAGTGTACTTGGAAAGAAAGCTGCAGCATTAACAG 2820  
Qy 2821 ATAGTTACTTGGGCTTTTGGCCATGGGAGAAAGAAAGTATTTTCCAGATCTTCTCGAGC 2880  
Db 2821 ATAGTTACTTGGGCTTTTGGCCATGGGAGAAAGAAAGTATTTTCCAGATCTTCTCGAGC 2880  
Qy 2881 AAGAGAGTCAATTTGAAGACACAAATTCGATACCTTCTGATAGCAAAATPACTGGGAGGC 2940  
Db 2881 AAGAGAGTCAATTTGAAGACACAAATTCGATACCTTCTGATAGCAAAATPACTGGGAGGC 2940  
Qy 2941 AACTAAAGATACATTTGAGATTCCTCAGATATGTAATAAATAAATTTCTAAATAGCAAGT 3000  
Db 2941 AACTAAAGATACATTTGAGATTCCTCAGATATGTAATAAATAAATTTCTAAATAGCAAGT 3000  
Qy 3001 TTGANTTACATCGCGAAAGTCCCTGCTCACATGCTCACATGATTTGACCGGATTTGTTA 3060

Db	3001	TTGGAATTCACATCGCGGAAAGTCCTGCTCATCATGCTCAGATGANTGACCGGATTTGTA	3060
Qy	3061	TGCAGAATCTCAAGATATGTTTCCTGAAGAATTTGACAAGACGTCAATTCACAAGATGC	3120
Db	3061	TGCAGAATCTCAAGATATGTTTCCTGAAGAATTTGACAAGACGTCAATTCACAAGATGC	3120
Qy	3121	GCCATTTGAGGATATGCAGTTTGCCTTCTCTTAATTTTAATTAATCTCATGAGTGCAGTGC	3180
Db	3121	GCCATTTGAGGATATGCAGTTTGCCTTCTCTTAATTTTAATTAATCTCATGAGTGCAGTGC	3180
Qy	3181	AGCCACTGAATATATCTCAAGCTCTTGATGAAGTTGATACAGATCAATCTGGTGCTTTGT	3240
Db	3181	AGCCACTGAATATATCTCAAGCTCTTGATGAAGTTGATACAGATCAATCTGGTGCTTTGT	3240
Qy	3241	CTGACAGAGAAATCCGAACACTGGCTACAGAAATTCACGAATCCGCCGTTAAGTTTGCAGG	3300
Db	3241	CTGACAGAGAAATCCGAACACTGGCTACAGAAATTCACGAATCCGCCGTTAAGTTTGCAGG	3300
Qy	3301	ATTTGACAGGTCGGAACACATGCTATAAATTCGTCAAAATAAGCTTCTCGTCGATATCA	3360
Db	3301	ATTTGACAGGTCGGAACACATGCTATAAATTCGTCAAAATAAGCTTCTCGTCGATATCA	3360
Qy	3361	CGCAGCTAAATAATATTCACCAACTCAGGAATCCTACTATGATCCCAAACTGCCACCGG	3420
Db	3361	CGCAGCTAAATAATATTCACCAACTCAGGAATCCTACTATGATCCCAAACTGCCACCGG	3420
Qy	3421	TCACTAAAAGTCTAGTAACAAACTGTAAACAGTAATCTGACAAAATCCACAAGCATATA	3480
Db	3421	TCACTAAAAGTCTAGTAACAAACTGTAAACAGTAATCTGACAAAATCCACAAGCATATA	3480
Qy	3481	AGGACAAAAACAATATAGGTTTGAAATCATGGGAGAGAAAGAAATCGCTTTTAAATATGA	3540
Db	3481	AGGACAAAAACAATATAGGTTTGAAATCATGGGAGAGAAAGAAATCGCTTTTAAATATGA	3540
Qy	3541	TTCTGACCAAGCTTTCTCATGTGTTGGCCAGTTGGATGACATAGAAGAAAAACCCTAGGA	3600
Db	3541	TTCTGACCAAGCTTTCTCATGTGTTGGCCAGTTGGATGACATAGAAGAAAAACCCTAGGA	3600
Qy	3601	AGTTTGTTTGCTGAAATGACAAACTTGACCAATCATTAAGATGCTCAGACAGTGAAGG	3660
Db	3601	AGTTTGTTTGCTGAAATGACAAACTTGACCAATCATTAAGATGCTCAGACAGTGAAGG	3660
Qy	3661	CTGTTCTCAGGGAATCTTATGAATCCATGTTCCCCATACCTTCCCAATTTGAACTGCCAA	3720
Db	3661	CTGTTCTCAGGGAATCTTATGAATCCATGTTCCCCATACCTTCCCAATTTGAACTGCCAA	3720
Qy	3721	GAGAGTATCGAAAACGGTTTCCTCATATGATGATGCTGCAGGAATGAGGGCTTATCGAG	3780
Db	3721	GAGAGTATCGAAAACGGTTTCCTCATATGATGATGCTGCAGGAATGAGGGCTTATCGAG	3780
Qy	3781	ACAAATTGAAGTTTGGACCCATTTGTACTAGCAACATTCATTATGTTTACTATATTCT	3840
Db	3781	ACAAATTGAAGTTTGGACCCATTTGTACTAGCAACATTCATTATGTTTACTATATTCT	3840
Qy	3841	CATTTTTTGTGAGAGTTAATTCGACTTAAGCGGAAAGATATTTCGCAGAGGAGGATAC	3900
Db	3841	CATTTTTTGTGAGAGTTAATTCGACTTAAGCGGAAAGATATTTCGCAGAGGAGGATAC	3900
Qy	3901	ACAAAGAGCTAGTCCCAATCGAATCAGAGTATAGAAGATCTTCATTTGAAAACCAATCTA	3960
Db	3901	ACAAAGAGCTAGTCCCAATCGAATCAGAGTATAGAAGATCTTCATTTGAAAACCAATCTA	3960
Qy	3961	CCTCAGCATTTACTGAGCATTTTAAAACTCAGCTTTCACAGAGATGTCTTTGTGATGTGAT	4020
Db	3961	CCTCAGCATTTACTGAGCATTTTAAAACTCAGCTTTCACAGAGATGTCTTTGTGATGTGAT	4020
Qy	4021	GCTTAGCAGTTTGGCCCGAAGAGGAAAATATCCAGTACCATGCTGTTTGTGGCATGAA	4080
Db	4021	GCTTAGCAGTTTGGCCCGAAGAGGAAAATATCCAGTACCATGCTGTTTGTGGCATGAA	4080
Qy	4081	TATAGCCCACTGACTAGGAATTTATTAACCAACCCTGAAAACCTGTGTGTCGAGCAGC	4140

4081	Db	TATAGCCCACTGACTAGAAATATTTAAACCAACCCACTGTAAAACTTGTGTGTGTCGAGCAGC	4141
4141	Qy	TCTGAACCTGATTTTACTTTTTAAAGAAATTTGCTCATGAGACCTGTCACTCTTTTTTATAAAAA	4201
4141	Db	TCTGAACTGATTTTACTTTTTAAAGAAATTTGCTCATGAGACCTGTCACTCTTTTTTATAAAAA	4201
4201	Qy	GGCTCACTGACAAGAGACAGCTGTTAATTTTCCCAACAGCAATCAATTGCAGACTAACTTTAT	4261
4201	Db	GGCTCACTGACAAGAGACAGCTGTTAATTTTCCCAACAGCAATCAATTGCAGACTAACTTTAT	4261
4261	Qy	TAGGAGAAGCCTATGCGCAGCTGGGAGTGATTTGCTAAGAGGCTCCAGTCTTTGCAATCCAA	4321
4261	Db	TAGGAGAAGCCTATGCGCAGCTGGGAGTGATTTGCTAAGAGGCTCCAGTCTTTGCAATCCAA	4321
4321	Qy	AGCCTTTTGCTAAAGTTTGACCTTTTTTTTTTTTTTTTTCATTTCCCATTTTAAAGTAGTTACTA	4381
4321	Db	AGCCTTTTGCTAAAGTTTGACCTTTTTTTTTTTTTTTTTCATTTCCCATTTTAAAGTAGTTACTA	4381
4381	Qy	AGTTAACTAGTTATCTTTGCTCTCGAGTATAACGAATTTGGGATGCTTAAACCTATTATTTA	4441
4381	Db	AGTTAACTAGTTATCTTTGCTCTCGAGTATAACGAATTTGGGATGCTTAAACCTATTATTTA	4441
4441	Qy	TAGATGTTATTTAAATTAATGCAACAATPACCTCTTTATTGACAATACTTAAATTTATGAG	4501
4441	Db	TAGATGTTATTTAAATTAATGCAACAATPACCTCTTTATTGACAATACTTAAATTTATGAG	4501
4501	Qy	TTTTTAAATTTAAAGACTGTAAATGGTCTTAAACCACTAACTACTGGAAGCTCAATG	4561
4501	Db	TTTTTAAATTTAAAGACTGTAAATGGTCTTAAACCACTAACTACTGGAAGCTCAATG	4561
4561	Qy	ATTGACATCTGAATGCTTTGTAAATTTGACITTCAGCCCTAAGAAATGCTATGATTTCA	4621
4561	Db	ATTGACATCTGAATGCTTTGTAAATTTGACITTCAGCCCTAAGAAATGCTATGATTTCA	4621
4621	Qy	CGTGCAGTCTAAATTTCAACAGGCTAGAGTTAGTACTACTTACAGATGTAATTTATGTTT	4681
4621	Db	CGTGCAGTCTAAATTTCAACAGGCTAGAGTTAGTACTACTTACAGATGTAATTTATGTTT	4681
4681	Qy	TGGAAATGTACATATTTCAAAACAGAAATGCTCTCAATTTAGAAATAGTAGTCTGATGGCA	4741
4681	Db	TGGAAATGTACATATTTCAACAGAAATGCTCTCAATTTAGAAATAGTAGTCTGATGGCA	4741
4741	Qy	CTGSCAATTAACAGTGTGTCTTTTAAATACATCATTTGATATATTCAGTAGCTATCTCT	4801
4741	Db	CTGSCAATTAACAGTGTGTCTTTTAAATACATCATTTGATATATTCAGTAGCTATCTCT	4801
4801	Qy	CTCAGTTGGTTTTGATAGAACAGAGCCAGCAAACTTTCTTTGTAAGGCTGGTTAGT	4861
4801	Db	CTCAGTTGGTTTTGATAGAACAGAGCCAGCAAACTTTCTTTGTAAGGCTGGTTAGT	4861
4861	Qy	AAATATTGAGGCCACCTGTGCTTTTGTGCATACATTTCTTTGCTGTTTGTAGTTGT	4921
4861	Db	AAATATTGAGGCCACCTGTGCTTTTGTGCATACATTTCTTTGCTGTTTGTAGTTGT	4921
4921	Qy	TTTTTTTTCAACAACCTCTAAAAATGTAATAACCATTTTACCTGTCGACTCTGACAAA	4981
4921	Db	TTTTTTTTCAACAACCTCTAAAAATGTAATAACCATTTTACCTGTCGACTCTGACAAA	4981
4981	Qy	ACTGCCACAGCCAGATGTGACCCCTCAGGCCATCATTTGCCAATCACTCAGAAATTTT	5041
4981	Db	ACTGCCACAGCCAGATGTGACCCCTCAGGCCATCATTTGCCAATCACTCAGAAATTTT	5041
5041	Qy	TTGTTGTTGTTGTTGTTGTTTTCAGACAGAGTCTCTCTCTGTTGCCAGCTGGAG	5101
5041	Db	TTGTTGTTGTTGTTGTTGTTTTCAGACAGAGTCTCTCTCTGTTGCCAGCTGGAG	5101
5101	Qy	TGCAGTGGCGCAATCTCAGCTCACTGCAACCTCCGGCTCCCGGGTTCAGACAGTCTCTG	5161
5101	Db	TGCAGTGGCGCAATCTCAGCTCACTGCAACCTCCGGCTCCCGGGTTCAGACAGTCTCTG	5161
5161	Qy	TCAGCCCTCTGAGTAGCTGGACTACAGGTGCAATGCCACACACCTCTGTAATTTTGTGA	5221
5161	Db	TCAGCCCTCTGAGTAGCTGGACTACAGGTGCAATGCCACACACCTCTGTAATTTTGTGA	5221





1381 ATGATGATGTCATGTTGGGAAGGATGTCGGCCAGATGATTTTACAGTCACTCCAAAG 1440  
1381 ATGATGATGTCATGTTGGGAAGGATGTCGGCCAGATGATTTTACAGTCACTCCAAAG 1440  
1441 GCCAAGAGGTTTATTTGACATGGCTGTGCCAAACTGTGCCAGGGCTGCCAGGTTTCCT 1500  
1441 GCCAAGAGGTTTATTTGACATGGCTGTGCCAAACTGTGCCAGGGCTGCCAGGTTTCCT 1500  
1501 GGATTAAGGATGGCTTATTTGCAAGGCTTGTATAAATTCAGCTGCGATGGAGTGGTG 1560  
1501 GGATTAAGGATGGCTTATTTGCAAGGCTTGTATAAATTCAGCTGCGATGGAGTGGTG 1560  
1561 GGGATGCTCTGGAAACAGTGGAGGAGTCCCTATATTTGCAAGGAGTGGAGGTTACTGGGA 1620  
1561 GGGATGCTCTGGAAACAGTGGAGGAGTCCCTATATTTGCAAGGAGTGGAGGTTACTGGGA 1620  
1621 GTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAAGTGTCTTACTGTGA 1680  
1621 GTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAAGTGTCTTACTGTGA 1680  
1681 ATCAGGAGTGGCGAATTCCTGGCTCGCTGATTAAGTCTGTGACCAAGCATGCAATGTCT 1740  
1681 ATCAGGAGTGGCGAATTCCTGGCTCGCTGATTAAGTCTGTGACCAAGCATGCAATGTCT 1740  
1741 TGTCTGTGGGTTTGATGCTGGCAGCTGTGGCAAGATCATTTTCATGAATTTGTATAAAG 1800  
1741 TGTCTGTGGGTTTGATGCTGGCAGCTGTGGCAAGATCATTTTCATGAATTTGTATAAAG 1800  
1801 TGATCTCTCCCAACCAAGATCTCATATATTTTCCAAAGAGTGAATGCGCTTATT 1860  
1801 TGATCTCTCCCAACCAAGATCTCATATATTTTCCAAAGAGTGAATGCGCTTATT 1860  
1861 TCAGCTTTGCAAGAGTACCAAAAGAGGAGTTGAAGGTGCTATAGTGAACAATCCATAA 1920  
1861 TCAGCTTTGCAAGAGTACCAAAAGAGGAGTTGAAGGTGCTATAGTGAACAATCCATAA 1920  
1921 TTCGATGCTCTTATTTGCCAACAGTGGAAAAACATCCACCTCATATGACAGTGGAA 1980  
1921 TTCGATGCTCTTATTTGCCAACAGTGGAAAAACATCCACCTCATATGACAGTGGAA 1980  
1981 TGAATGCCACACATATATTTTATCTCAGCTTTCAAAATACAAAGATGAGAGTTCA 2040  
1981 TGAATGCCACACATATATTTTATCTCAGCTTTCAAAATACAAAGATGAGAGTTCA 2040  
2041 AAATCAGATTAACAGTGGAGTGGACACAGGGAGGAGCCAAAACCTGAAATCTACGGCCC 2100  
2041 AAATCAGATTAACAGTGGAGTGGACACAGGGAGGAGCCAAAACCTGAAATCTACGGCCC 2100  
2101 AGAAGGTTACGAAATTTAGTTAGTCCCATTAACATCTTCTCAGAGCGGAAATCTTTT 2160  
2101 AGAAGGTTACGAAATTTAGTTAGTCCCATTAACATCTTCTCAGAGCGGAAATCTTTT 2160  
2161 TTGAGCATATTTCCCAAGAAAAACGCTTCCCGAAGTTTAAAGACATGATTTAACTCAA 2220  
2161 TTGAGCATATTTCCCAAGAAAAACGCTTCCCGAAGTTTAAAGACATGATTTAACTCAA 2220  
2221 CAAGAGAGCCGAGGAGGTAATAATTCCTCTGGTAAATATTTTCTCTCCCAAG 2280  
2221 CAAGAGAGCCGAGGAGGTAATAATTCCTCTGGTAAATATTTTCTCTCCCAAG 2280  
2281 ACGCCAGTGTAGTCTCAATACCTTGGATTTGCAACTGGAAATGAGAGATCACTTTGA 2340  
2281 ACGCCAGTGTAGTCTCAATACCTTGGATTTGCAACTGGAAATGAGAGATCACTTTGA 2340  
2341 AAGGATCAATTTGTCCAAAGTCAAGCTTGTGAGATCATTTCTGATGAATTCAGCATG 2400  
2341 AAGGATCAATTTGTCCAAAGTCAAGCTTGTGAGATCATTTCTGATGAATTCAGCATG 2400  
2401 CTAAATTAATAATCAAGCTTAATAACAGATGAACAAATGACAGTTTGGTGGCTCCAC 2460  
2401 CTAAATTAATAATCAAGCTTAATAACAGATGAACAAATGACAGTTTGGTGGCTCCAC 2460  
2461 AGGAAAAACAGGTTTCATATAAAGCATCTTTGCCAAACAGCTTAGGAGTGTCTGAAAGATTGC 2520

2461 AGGAAAAACAGGTTTCATATAAAGCATCTTGCCAAACAGCTTAGGAGTGTCTGAAAGATTGC 2520  
2521 AGAGGTTGATCTTTCTCGAGTGTAAAAGTGAATGGTCAACAGGAGTGTGAGATC 2580  
2521 AGAGGTTGATCTTTCTCGAGTGTAAAAGTGAATGGTCAACAGGAGTGTGAGATC 2580  
2581 CACCCCTGGACTTGGAGACCAAGATTTAGAGTGGAAACTCACACCAAAACCA 2640  
2581 CACCCCTGGACTTGGAGACCAAGATTTAGAGTGGAAACTCACACCAAAACCA 2640  
2641 TAGGCGGAAATGTGACAAAAGAAAGCCCATCTCTGATTTGTTCCACTGGAAAGCCAGA 2700  
2641 TAGGCGGAAATGTGACAAAAGAAAGCCCATCTCTGATTTGTTCCACTGGAAAGCCAGA 2700  
2701 TGACAAAAGAAAGAAATTCACAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2760  
2701 TGACAAAAGAAAGAAATTCACAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2760  
2761 CTGAAATTCACATAGCGCTTACTGAAAGTGTACTTTGGAAAGAAAGCTGCAGCATTCACAG 2820  
2761 CTGAAATTCACATAGCGCTTACTGAAAGTGTACTTTGGAAAGAAAGCTGCAGCATTCACAG 2820  
2821 ATAGTTACTTGGCTTTTGGCATGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2880  
2821 ATAGTTACTTGGCTTTTGGCATGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2880  
2881 AAGAGAGTCAATTTGAGAGCAATTTGGCATCTTCACTGATAGCAAAATCTGGAGGC 2940  
2881 AAGAGAGTCAATTTGAGAGCAATTTGGCATCTTCACTGATAGCAAAATCTGGAGGC 2940  
2941 AACTAAAAGATACATTTGAGAGTCCCTCAGATATGTAATAAATAATCTAAATAGCAAGT 3000  
2941 AACTAAAAGATACATTTGAGAGTCCCTCAGATATGTAATAAATAATCTAAATAGCAAGT 3000  
3001 TTGGATTCACATCGGGAAGTCCCTGCTCAGTCCCTCAGATATGTAATAAATAATCTAAATAGCAAGT 3060  
3001 TTGGATTCACATCGGGAAGTCCCTGCTCAGTCCCTCAGATATGTAATAAATAATCTAAATAGCAAGT 3060  
3061 TGCAAGAACTGCAAGATATGTTCCCTGAAAGTAAATTTGCAAGAGCTCAATTTCAAAAGTGC 3120  
3061 TGCAAGAACTGCAAGATATGTTCCCTGAAAGTAAATTTGCAAGAGCTCAATTTCAAAAGTGC 3120  
3121 GGCATTTGAGATATGAGTTTGGCTTCTCTTATTTTATTTATCTCATGAGTGCAGTGC 3180  
3121 GGCATTTGAGATATGAGTTTGGCTTCTCTTATTTTATTTATCTCATGAGTGCAGTGC 3180  
3181 AGCCATGATATATCTCAAGTCTTTGATGAGTTGATACAGATCAATCTGGTGTCTTGT 3240  
3181 AGCCATGATATATCTCAAGTCTTTGATGAGTTGATACAGATCAATCTGGTGTCTTGT 3240  
3241 CTGACAGAAATCCGAAACACCTGGCTACAGAAATTCAGAACTGCGGTTAAATTTGCAAG 3300  
3241 CTGACAGAAATCCGAAACACCTGGCTACAGAAATTCAGAACTGCGGTTAAATTTGCAAG 3300  
3301 ATTTGACAGCTGCGAAACACATGCTAAATTTGCTCAAAATGCTTCTGCTGATATCA 3360  
3301 ATTTGACAGCTGCGAAACACATGCTAAATTTGCTCAAAATGCTTCTGCTGATATCA 3360  
3361 CGCAGCTAAATATTTCCACCACTCAGGAATCTTACTATGATCCCACTGCCACCGG 3420  
3361 CGCAGCTAAATATTTCCACCACTCAGGAATCTTACTATGATCCCACTGCCACCGG 3420  
3421 TCATTAAGAGTCTAGTAAACACTGTAAACAGTAACTGACAAATCCCAAGCATATA 3480  
3421 TCATTAAGAGTCTAGTAAACACTGTAAACAGTAACTGACAAATCCCAAGCATATA 3480  
3481 AGGACAAAACAAATATAGGTTTGAATCATGGGAGAAAGAAATCGCTTTTAAATGA 3540  
3481 AGGACAAAACAAATATAGGTTTGAATCATGGGAGAAAGAAATCGCTTTTAAATGA 3540  
3541 TTCGTACCAAGCTTCTCATGTTGGCTGAGTGGAGTGGATGACATAGAAAGAAACCTTAGGA 3600



3541 TTCTGACCAACGTTTCTCATGTGGTTGGCCAGTTGGATGACATGAAGAAAAACCCCTAGGA 3600  
3601 AGTTTGTGTGCTGAATGACAAACATTTGACCAACATCATATAAGATGCTCAGACAGTGAAG 3660  
3601 AGTTTGTGTGCTGAATGACAAACATTTGACCAACATCATATAAGATGCTCAGACAGTGAAG 3660  
3661 CTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAACTGCCAA 3720  
3661 CTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAACTGCCAA 3720  
3721 GAGATATCGAAACCGTTTCTTCATATGATGATGATGATGATGATGATGATGATGATGATGAT 3780  
3721 GAGATATCGAAACCGTTTCTTCATATGATGATGATGATGATGATGATGATGATGATGATGAT 3780  
3781 ACAAAATCGAATTTTGGACCAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 3840  
3781 ACAAAATCGAATTTTGGACCAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 3840  
3841 CATTTTTTCTCAGCAGTTAAATGACATTAAGCGGAAGATATTTCCAGAGAGGAGATAC 3900  
3841 CATTTTTTCTCAGCAGTTAAATGACATTAAGCGGAAGATATTTCCAGAGAGGAGATAC 3900  
3901 ACAAGAGAGTATGCTCCCAATCGAATCAGAGTATAGAAGATCTTCAATTTGAAACCAATCTA 3960  
3901 ACAAGAGAGTATGCTCCCAATCGAATCAGAGTATAGAAGATCTTCAATTTGAAACCAATCTA 3960  
3961 CCTCAGCATTTACTCAGCATTTTAAATCAGCTTACAGAGATGCTCTTTGTGATGATGAT 4020  
3961 CCTCAGCATTTACTCAGCATTTTAAATCAGCTTACAGAGATGCTCTTTGTGATGATGAT 4020  
4021 GCTTACAGTGTGGCCCGAAGAAATATCCAGTACCAATGCTGTTTGTGATGATGATGATGAT 4080  
4021 GCTTACAGTGTGGCCCGAAGAAATATCCAGTACCAATGCTGTTTGTGATGATGATGATGAT 4080  
4081 TATAGCCCACTGACTAGGAATTTTAAACCAACCACTGAAACCTTGTGCTGCGAGCAGC 4140  
4081 TATAGCCCACTGACTAGGAATTTTAAACCAACCACTGAAACCTTGTGCTGCGAGCAGC 4140  
4141 TCTGAACCTGATTTTAAAGAAATTTGCTATGACCTGCTCATCTCTTTTATAAAAA 4200  
4141 TCTGAACCTGATTTTAAAGAAATTTGCTATGACCTGCTCATCTCTTTTATAAAAA 4200  
4201 GCTCAGTACGACAGAGAGCTGTTAAATTTCCAGAGCAATCATTTGAGAGCTTAT 4260  
4201 GCTCAGTACGACAGAGAGCTGTTAAATTTCCAGAGCAATCATTTGAGAGCTTAT 4260  
4261 TAGGAGAGCTTATCCAGCTGGAGTGAATTAAGAGGCTCCAGTCTTTGCAATCCAA 4320  
4261 TAGGAGAGCTTATCCAGCTGGAGTGAATTAAGAGGCTCCAGTCTTTGCAATCCAA 4320  
4321 AGCTTTTCTGCTAAAGTGTGCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4380  
4321 AGCTTTTCTGCTAAAGTGTGCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4380  
4381 AGTTAACTAGTTATTTCTGCTTCTGATTAACGAATTTGGGATGCTTAAACCTATTTT 4440  
4381 AGTTAACTAGTTATTTCTGCTTCTGATTAACGAATTTGGGATGCTTAAACCTATTTT 4440  
4441 TAGATGTTATTTAAATAATGACAGCAATATCACCTCTTATGACAAATACCTAAATATGAG 4500  
4441 TAGATGTTATTTAAATAATGACAGCAATATCACCTCTTATGACAAATACCTAAATATGAG 4500  
4501 TTTTATTAATAATTTAAGACTGAATGCTTTAAACCACTTAACCTAATGAGAGCTCAATG 4560  
4501 TTTTATTAATAATTTAAGACTGAATGCTTTAAACCACTTAACCTAATGAGAGCTCAATG 4560  
4561 ATTGACATCTGAATGCTTTTAAATTTATTTGCTTCCAGCCCTTAAGAAATGCTATGATTTCA 4620  
4561 ATTGACATCTGAATGCTTTTAAATTTATTTGCTTCCAGCCCTTAAGAAATGCTATGATTTCA 4620  
4621 CTTGAGGCTTAAATTTCAACAGGCTAGAGTTAGTACTACTTACCAAGATGTAATATGTTT 4680  
4621 CTTGAGGCTTAAATTTCAACAGGCTAGAGTTAGTACTACTTACCAAGATGTAATATGTTT 4680

4681 TGGAAATGATACATATTCAAAACAGAGTGCTCTCATTTTAGAAATAGTAGTGTCTGATGGCA 4740  
4681 TGGAAATGATACATATTCAAAACAGAGTGCTCTCATTTTAGAAATAGTAGTGTCTGATGGCA 4740  
4741 CTGCAATATACAGTGGTGTCTTTTAATACATCTCATTTGGTATATTCAGTAGTCTATCTCT 4800  
4741 CTGCAATATACAGTGGTGTCTTTTAATACATCTCATTTGGTATATTCAGTAGTCTATCTCT 4800  
4801 CTGAGTTGGTTTTTGTATAGAAACAGAGCCAGCAAACTTTCTTTGTAAGAGGCTGGTTAGT 4860  
4801 CTGAGTTGGTTTTTGTATAGAAACAGAGCCAGCAAACTTTCTTTGTAAGAGGCTGGTTAGT 4860  
4861 AAATATTGAGGCCACCTGTGTCTTTGTCATACATCTCTCTCTGCTGTTGTTGTTGTTGTTG 4920  
4861 AAATATTGAGGCCACCTGTGTCTTTGTCATACATCTCTCTCTGCTGTTGTTGTTGTTGTTG 4920  
4921 TTTTTCCTCAAAACCCCTCTTAAATATGTAACCAATGTTAGCTTGCAGCTGTACAAA 4980  
4921 TTTTTCCTCAAAACCCCTCTTAAATATGTAACCAATGTTAGCTTGCAGCTGTACAAA 4980  
4981 ACTGCCACACAGCAGATGTCAGCCCTCAGCCCATCATTTGCCAATCAGTGAAGATTTT 5040  
4981 ACTGCCACACAGCAGATGTCAGCCCTCAGCCCATCATTTGCCAATCAGTGAAGATTTT 5040  
5041 TTG 5100  
5041 TTG 5100  
5101 TGAGTGGCCCAATCTCAGCTCACTGCAACCTCCGCTCCCGGTTCAAGCAGTCTGTC 5160  
5101 TGAGTGGCCCAATCTCAGCTCACTGCAACCTCCGCTCCCGGTTCAAGCAGTCTGTC 5160  
5161 TCAGCTTCTGAGTGTGGGACTTACAGGTGTCATGCCACACACCTGCTTAATTTTGTGA 5220  
5161 TCAGCTTCTGAGTGTGGGACTTACAGGTGTCATGCCACACACCTGCTTAATTTTGTGA 5220  
5221 TTTTATGATAGACGGGGTTCCACCATATGCTCAGGCTTATCTTGAACCTCTGACCTC 5280  
5221 TTTTATGATAGACGGGGTTCCACCATATGCTCAGGCTTATCTTGAACCTCTGACCTC 5280  
5281 AGGTGATCCACCTGCTCTGCTCCCAAGTGTGAGATTACAGGATAGCCAGTGCAC 5340  
5281 AGGTGATCCACCTGCTCTGCTCCCAAGTGTGAGATTACAGGATAGCCAGTGCAC 5340  
5341 CCAGCCGAGAAATTTATGTTTATGATGTTAAACCTTTGGGCTCTAGCCATATTTATG 5400  
5341 CCAGCCGAGAAATTTATGTTTATGATGTTAAACCTTTGGGCTCTAGCCATATTTATG 5400  
5401 TCATAATACATGATTTGTGAAGACAGATTCATGAGTACTCTGACAGTATTTTAG 5460  
5401 TCATAATACATGATTTGTGAAGACAGATTCATGAGTACTCTGACAGTATTTTAG 5460  
5461 ATCATGATCTCAACAATATTCCTCCCAATGGCATACATCTTTTGTACAAAGAACTTGAA 5520  
5461 ATCATGATCTCAACAATATTCCTCCCAATGGCATACATCTTTTGTACAAAGAACTTGAA 5520  
5521 ATGTAATACATGTTGTTGCTGTAGAGTGTGTTATTTCAAAATCTGAAATCTCATAA 5580  
5521 ATGTAATACATGTTGTTGCTGTAGAGTGTGTTATTTCAAAATCTGAAATCTCATAA 5580  
5581 AAGTTAAATTTTCAAAA 5597  
5581 AAGTTAAATTTTCAAAA 5597

RESULT 4  
AB033034  
LOCUS Homo sapiens mRNA for KIAA1208 protein, partial cds.  
DEFINITION  
AB033034  
ACCESSION  
VERSION  
AB033034.2 GI:20521791  
KEYWORDS

SOURCE	Homo sapiens (human)	Db	121	TTTCATTGTACCAACGGGAGATTCATCTCTGGCTGAACCTTGACATCTCTCGAGTAC	180
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	Qy	1262	AAATAGTAACACACCGAGATGTTTTTCGAAATTTAGAGCCATCTGCTTACCTTACCTACC	1321
AUTHORS	1 Nagase,T., Ishikawa,K., Kikuno,R., Hirose,M., Nomura,N. and Ohara,O.	Db	181	AAATAGTAACACACCGAGATGTTTTTCGAAATTTAGAGCCATCTGCTTACCTTACCTACC	240
TITLE	Prediction of the coding sequences of unidentified human genes. XV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro	Qy	1322	TGCTATTGAAAGTCACATTCATCGCATGAAGGGCTGCCAGAGATTTTATTTACTATAA	1381
JOURNAL	DNA Res. 6 (5), 337-345 (1999)	Db	241	TGCTATTGAAAGTCACATTCATCGCATGAAGGGCTGCCAGAGATTTTATTTACTATAA	300
PUBMED	10574462				
REFERENCE	2 (bases 1 to 4511)	Qy	1382	TGATGATGTCATGTTTGGGAAGGATGTCTGGCCAGATGATTTTACAGTCACTCCAAAGG	1441
AUTHORS	Ohara,O., Nagase,T. and Kikuno,R.	Db	301	TGATGATGTCATGTTTGGGAAGGATGTCTGGCCAGATGATTTTACAGTCACTCCAAAGG	360
TITLE	Direct Submission	Qy	1442	CCAGAAGGTTTATTTGACATGCTGTCACCAAGTGTCCGAGGGCTGCCAGGTTCTCTG	1501
JOURNAL	Submitted (04-OCT-1999) Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology; 1532-3 Yana, Kisarazu, Chiba	Db	361	CCAGAAGGTTTATTTGACATGCTGTCACCAAGTGTCCGAGGGCTGCCAGGTTCTCTG	420
REFERENCE	292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, URL:http://www.kazusa.or.jp/huge/, Tel:+81-438-52-3913, Fax:+81-438-52-3914)	Qy	1502	GATTAAAGGATGCTTATTTGACAGGCTTGTAAATTAATTCAGCCTCGGATTTGGGATGGG	1561
COMMENT	On May 9, 2002 this sequence version replaced gi:6382021.	Db	421	GATTAAAGGATGCTTATTTGACAGGCTTGTAAATTAATTCAGCCTCGGATTTGGGATGGG	480
FEATURES	Location/Qualifiers				
source	1..4511	Qy	1562	GGATTGCTCTGGAACACAGTGGAGGGAGTGCCTATATTTGACAGGAGTGGAGTACCTGGAG	1621
gene	/organism="Homo sapiens"	Db	481	GGATTGCTCTGGAACACAGTGGAGGGAGTGCCTATATTTGACAGGAGTGGAGTACCTGGAG	540
CDS	/mol_type="mRNA"	Qy	1622	TATTGAGTGGAGACACCTCTGCGAGTGTGGTGGAGGATTAACACAGTGTCTCTTACTGTAA	1681
	/db_xref="taxon:9606"	Db	541	TATTGAGTGGAGACACCTCTGCGAGTGTGGTGGAGGATTAACACAGTGTCTCTTACTGTAA	600
	/clone="fj07955"	Qy	1682	TCAGGAGTGGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1741
	/tissue_type="brain"	Db	601	TCAGGAGTGGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	660
	/clone_lib="pbluescriptII SK plus"	Qy	1742	GTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1801
	/note="This sequence was replaced with that of f095318 cDNA as a representative cDNA sequence for KIAA1208."	Db	661	GTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	720
	1..4511	Qy	1802	GATCTCTCTCCCAACACAGACTCCTATATTTATTTCCAAAAGGTGAATGCCCTTATTT	1861
	/gene="KIAA1208"	Db	721	GATCTCTCTCCCAACACAGACTCCTATATTTATTTCCAAAAGGTGAATGCCCTTATTT	780
	<1..2854	Qy	1862	CAGCTTTGCAAGTAGCCCAAGAGGAGTTCGAAGTGCCTTATAGTGCATTCCTCAATAAT	1921
	/gene="KIAA1208"	Db	781	CAGCTTTGCAAGTAGCCCAAGAGGAGTTCGAAGTGCCTTATAGTGCATTCCTCAATAAT	840
	/note="Start codon is not identified."	Qy	1922	TGCACATGCTCTTATTTGCCAACACAGTGGAAACCATCCACCTCATATAATGCAAGTGAAT	1981
	/product="KIAA1208 protein"	Db	841	TGCACATGCTCTTATTTGCCAACACAGTGGAAACCATCCACCTCATATAATGCAAGTGAAT	900
	/protein_id="BAAB6522.2"	Qy	1982	GAATGCCACCAATATCTTTTAAATCTCAAGTTCCTCAATAATGCAAGTGAATGCTCAAT	2041
	/db_xref="GI:20521792"	Db	901	GAATGCCACCAATATCTTTTAAATCTCAAGTTCCTCAATAATGCAAGTGAATGCTCAAT	960
	/translation="SAISQSKQDEDISARFEDNEELYSLSIERHAPWVRNIFIVT NGQIPSLNLDNPVITVTHQVYLPVNCAGCGPGWIKGDCDRAACNSADWGG VMDKDVDFDYSHGKQKQYLVFVNCAGCGPGWIKGDCDRAACNSADWGG DCSGNSGSRVYGGGIGVQWPGQVGGINSVYCNQGCANSLWADKCDQCN VLSGCFDAGCGQDHFHLYKVLTPQTHYIIPKGECLPFYFAEYKRGVEGAYSD NPTRHASIANKNTIHLIHSGMNTIHFNTLQNTNDEEFMQITVVEVDTRGPK LNSTACQGVNLVSPITLLPBAELTFEDIPEKRPFPKRDHVNSTRARQEVKILIV NISLLPKDAQLSLNLDLQLEPHGDI TLKYNLSKALLRSLFNNSHAKIKNOAIIID ETNSLVAQKQVHKSLNLSGVSRERLQRLFPVAVSVKVGHDQGNPPLDLETA RFRVETHQITGNVTKPKPPSLIVPLESOMTKERKITGKEKNEENAEHNLGV TEVLGLKQHTSYSLGFLPWEKKYFQDLDBEESLKTOLAYFTDSKNTGROLKDT PADSLRYKNTLNSKFGTERKVPAPHPMIDRIVMOELQMPFEEDKTSFKVRS EDMQFARSFYFLYMSAVQPLNISQVDEVDTQSGVLSDEIRTLATRIHELPLSQD LTGLEHMLNCSKMLPADITOLNIPPTQSYVDPNLPVPTKSLVTKCKPVTDKIHKA YKDNKVFETWEGEEIAFKMRTNVSHVVGOLDIRKNRPKTVCLNDINDHNDKDAQ TVKAVLRDFYSMPFIQSQELPREYRNLHHELEQWEPAYRDKLKFTHCHVLAATLI MFTIFSFAEOLIAKPKIPFRRIHKEASPNRIRV"	Qy	2042	AATGACAGATAACAGTGGAGGTGGACACAGGAGGAGGACCAAACTGAATTTCTACGGCCCA	2101
		Db	961	AATGACAGATAACAGTGGAGGTGGACACAGGAGGAGGACCAAACTGAATTTCTACGGCCCA	1020
ORIGIN		Qy	2102	GAAGGTTTACGAAATTTAGTTAGTCCCATTAACATTTCTTCCAGAGGGGAAATCTTTTT	2161
Query Match	80.0%; Score 4475.2; DB 9; Length 4511;	Db	1021	GAAGGTTTACGAAATTTAGTTAGTCCCATTAACATTTCTTCCAGAGGGGAAATCTTTTT	1080
Best Local Similarity	99.8%; Pred. No. 0;	Qy	2162	TCAGGATATTCCTCAAGAAACCGCTTCCGAGGTTTAAAGACATGATGTTTAACTCAAC	2221
Mismatches 4502; Conservative	0; Mismatches 8; Indels 2; Gaps 2;	Db	1081	TCAGGATATTCCTCAAGAAACCGCTTCCGAGGTTTAAAGACATGATGTTTAACTCAAC	1140
Qy	1082 GAGGCCCATGACCGAGTAAAGAGGATGAAGACATCTCTGCGAGTGGTCTTGAAGATAA	1141			
Db	1 GAGGCCCATGACCGAGTAAAGAGGATGAAGACATCTCTGCGAGTGGTCTTGAAGATAA	60			
Qy	1142 CGAAGAACTGAGTACTCATTCGCGATCTATCGAGGAGTGCACCATGGTTCGGAATAT	1201			
Db	61 CGAAGAACTGAGTACTCATTCGCGATCTATCGAGGAGTGCACCATGGTTCGGAATAT	120			
Qy	1202 TTTCATTGTACCAACGGGAGATTCATCTCTGGCTGAACCTTGACATCTCTCGAGTAC	1261			

100

...  
...  
...  
...  
...  
...  
...  
...

3



1652 TGGAGGAATAAACAGTGTCTCTTACTGTAAATCAGGATGTGGAATTCCTGGCTCGCTGA 1711  
1621 TGGAGGAATAAACAGTGTCTCTTACTGTAAATCAGGATGTGGAATTCCTGGCTCGCTGA 1680  
1712 TAAAGTCTGTGACCAAGCATGCAATGCTGTGCTCTGTGGTTTGTAGTGTGGGACCTGTGG 1771  
1681 TAAAGTCTGTGACCAAGCATGCAATGCTGTGCTCTGTGGTTTGTAGTGTGGGACCTGTGG 1740  
1772 GCAAGATCATTTTCATGAATTTGATAAAGTGTATCCCTTCCCAAAACACAGACTCACATAT 1831  
1741 GCAAGATCATTTTCATGAATTTGATAAAGTGTATCCCTTCCCAAAACACAGACTCACATAT 1800  
1832 TATTCCAAAGGTGAATGCTGCTTTATTTCAGCTTTTGAGAGTGTAGCCAAAGAGGAGT 1891  
1801 TATTCCAAAGGTGAATGCTGCTTTATTTCAGCTTTTGAGAGTGTAGCCAAAGAGGAGT 1860  
1892 TGAAGTGTCTTGTAGTACCAATTCCTTCTGACATGCTTCTTATTCCTGCAACAGTGGAA 1951  
1861 TGAAGTGTCTTGTAGTACCAATTCCTTCTGACATGCTTCTTATTCCTGCAACAGTGGAA 1920  
1952 AACCATCCCTCATTAATGCACAGTGAATGCAATGCACCAACACATATATTTTAACTCTAC 2011  
1921 AACCATCCCTCATTAATGCACAGTGAATGCAATGCACCAACACATATATTTTAACTCTAC 1980  
2012 GTTTCAAAATAAAGATGAGAGTTCAAAATGACAGATAACAGTGGAGGTGACACACAG 2071  
1981 GTTTCAAAATAAAGATGAGAGTTCAAAATGACAGATAACAGTGGAGGTGACACACAG 2040  
2072 GGAAGGACCAAACTGAAATTTACGGCCAGAGGCTTACGAAATTTAGTTAGTCCCAT 2131  
2041 GGAAGGACCAAACTGAAATTTACGGCCAGAGGCTTACGAAATTTAGTTAGTCCCAT 2100  
2132 AACCATCTTCCAGAGCGGAAATCTTTTGTAGGATATTCCTTCCAAAGAAACAGCTTCCC 2191  
2101 AACCATCTTCCAGAGCGGAAATCTTTTGTAGGATATTCCTTCCAAAGAAACAGCTTCCC 2160  
2192 GAAGTTTAAAGACATGATGTTAACTCAACAGGAGAGCCAGAGAGAGTGAATTC 2251  
2161 GAAGTTTAAAGACATGATGTTAACTCAACAGGAGAGCCAGAGAGAGTGAATTC 2220  
2252 CCTGGTAAATATTTTCACTCCTTCCAAAGACGCTTGTAGTCTCAATACCTTGGATT 2311  
2221 CCTGGTAAATATTTTCACTCCTTCCAAAGACGCTTGTAGTCTCAATACCTTGGATT 2280  
2312 GCACTTGAACATGAGAGATCATCTTTGAAAGATACAATTTGTCCAAGTCAAGCTTGTCT 2371  
2281 GCACTTGAACATGAGAGATCATCTTTGAAAGATACAATTTGTCCAAGTCAAGCTTGTCT 2340  
2372 GAGATCATTTCTGATGAATCTCAGCATGCTTAAATATAAATAAATCAAGCTATAATAACAG 2431  
2341 GAGATCATTTCTGATGAATCTCAGCATGCTTAAATATAAATAAATCAAGCTATAATAACAG 2400  
2432 TGAACCAATGACAGTTTGGTGGCTCCACAGGAAAAACAGGTTTCAATAAAGCATCTTGCC 2491  
2401 TGAACCAATGACAGTTTGGTGGCTCCACAGGAAAAACAGGTTTCAATAAAGCATCTTGCC 2460  
2492 AAACAGCTTAGGAGTGTCTGAAAGATTGCAAGAGTTTGAATTTCTTCTGAGTGTGATAA 2551  
2461 AAACAGCTTAGGAGTGTCTGAAAGATTGCAAGAGTTTGAATTTCTTCTGAGTGTGATAA 2520  
2552 AGTGAATGGTTCATGACAGGCTCAGAAATCCACCTTGGACTTGGAGACCAAGAGATT 2611  
2521 AGTGAATGGTTCATGACAGGCTCAGAAATCCACCTTGGACTTGGAGACCAAGAGATT 2580  
2612 TAGAGTGGAACTCAGACCCCAAAACCATAGCGGAAATGTGACAAAAGAAAGCCCC 2671  
2581 TAGAGTGGAACTCAGACCCCAAAACCATAGCGGAAATGTGACAAAAGAAAGCCCC 2640  
2672 ATCTCTGATTTTCCACTTGGAAAGCCAGATGACAAAAGAAAGAAATCACAGGGAAGA 2731  
2641 ATCTCTGATTTTCCACTTGGAAAGCCAGATGACAAAAGAAAGAAATCACAGGGAAGA 2700  
2732 AAAAGAAACAGTAGAATGGAGGAAATGCTGAAATACATAGGCTTACTAGGAGTGT 2791

572 GCCAATGCTGTACTGACCCAGCCCTGCCAGCAACATCACCTGTAGGAGCTGGCCATC 631  
541 GCCAATGCTGTACTGACCCAGCCCTGCCAGCAACATCACCTGTAGGAGCTGGCCATC 600  
632 TCTTTATCTCTTTTCAATTTCTGCCAGTGACATTTTCAATGTTGCAAAACCAAAACCC 691  
601 TCTTTATCTCTTTTCAATTTCTGCCAGTGACATTTTCAATGTTGCAAAACCAAAACCC 660  
632 TTCTACCAATGCTCAGTGTGTTTGTGACAGTACTAAGATGTTGAAGTGGCCACTC 751  
661 TTCTACCAATGCTCAGTGTGTTTGTGACAGTACTAAGATGTTGAAGTGGCCACTC 720  
752 TGGAGTGTCTTAAAGGAATAGCAGACAGACAGTATGAGAGGGGTACTTGACAAACAGATAA 811  
721 TGGAGTGTCTTAAAGGAATAGCAGACAGACAGTATGAGAGGGGTACTTGACAAACAGATAA 780  
812 AGAAGTCTCTGATTTAGTGTGTTTGTGACAGTACTAAGATGTTGAAGTGGCCACTC 871  
781 AGAAGTCTCTGATTTAGTGTGTTTGTGACAGTACTAAGATGTTGAAGTGGCCACTC 840  
872 ATTCAGGAAACCAATCAACTTAAACCAAAATGCGCAAAATCTTCTCTTAAAGTCAA 931  
841 ATTCAGGAAACCAATCAACTTAAACCAAAATGCGCAAAATCTTCTCTTAAAGTCAA 900  
932 ACTGTTGAGTGTGATTCAGAGGCCAGTGTAGCGCTTCTTAAACTGAAATAACCCCAAGGA 991  
901 ACTGTTGAGTGTGATTCAGAGGCCAGTGTAGCGCTTCTTAAACTGAAATAACCCCAAGGA 960  
992 TTTTCAAGATTTGAATAGCAACTAAGAGACATGACCATGATGAGAAAGAACTGAC 1051  
961 TTTTCAAGATTTGAATAGCAACTAAGAGACATGACCATGATGAGAAAGAACTGAC 1020  
1052 CATAGTCTCTGATATTTATTTATGGAATCTGAGCGCCATCAGGCAATCTAAGCAGGATGA 1111  
1021 CATAGTCTCTGATATTTATTTATGGAATCTGAGCGCCATCAGGCAATCTAAGCAGGATGA 1080  
1112 AGACATCTCTGCGAGTGTGTTGAAGATGAACGAGAACTGAGTACTCATTTGCGATCTAT 1171  
1081 AGACATCTCTGCGAGTGTGTTGAAGATGAACGAGAACTGAGTACTCATTTGCGATCTAT 1140  
1172 CGAGAGGATGACCAATGGTTTCGGAATATTTTCAATGTCACCAACGGGAGATTTCCATC 1231  
1141 CGAGAGGATGACCAATGGTTTCGGAATATTTTCAATGTCACCAACGGGAGATTTCCATC 1200  
1232 CTGGCTGAACCTTGAACAATCTCTGAGTGAACAATAGTAAACACCAAGGATGTTTTCGAAA 1291  
1201 CTGGCTGAACCTTGAACAATCTCTGAGTGAACAATAGTAAACACCAAGGATGTTTTCGAAA 1260  
1292 TTTGAGCCACTTGCCTACCTTTAGTTCACCTGCTATTGAAAGTCAATTCATTCGATCGA 1351  
1261 TTTGAGCCACTTGCCTACCTTTAGTTCACCTGCTATTGAAAGTCAATTCATTCGATCGA 1320  
1352 AGGGCTGTCCAGAAAGTTTATTTACCTAAATGATGATGTCATGTTTGGGAAAGGATGCTG 1411  
1321 AGGGCTGTCCAGAAAGTTTATTTACCTAAATGATGATGTCATGTTTGGGAAAGGATGCTG 1380  
1412 GCCAGATGATTTTTCAGTCACTCCAAAGGCCAGAGGTTTATTTGACATGGCTTGCC 1471  
1381 GCCAGATGATTTTTCAGTCACTCCAAAGGCCAGAGGTTTATTTGACATGGCTTGCC 1440  
1472 AAACCTGTGCGGAGGCTGCGCAGGTTCTTGGATTAAGGATGCTTATTTGCAAGGCTTG 1531  
1441 AAACCTGTGCGGAGGCTGCGCAGGTTCTTGGATTAAGGATGCTTATTTGCAAGGCTTG 1500  
1532 TAAATAATTCAGCTCGGATTTGGATGGTGGGATTTGCTCTGAAACAGTGGAGGAGTCG 1591  
1501 TAAATAATTCAGCTCGGATTTGGATGGTGGGATTTGCTCTGAAACAGTGGAGGAGTCG 1560  
1592 CTATATTCAGAGGTTGAGGTTACTGGGATTTGGAGTTGACACCCCTGCGAGTTTGG 1651  
1561 CTATATTCAGAGGTTGAGGTTACTGGGATTTGGAGTTTGGACAGCCCTGCGAGTTTGG 1620



2701 AAAAGAGAACAGTAGAATGGAGGAAAATGCTGAAAATCAGATAGGCGTTACTGAAGTGT 2760  
2792 ACTTGGAGAGAAAGCTGCAGCAATACACAGATAGTACTTGGGCTTTTGGCATGGAGAA 2851  
2761 ACTTGGAGAGAAAGCTGCAGCAATACACAGATAGTACTTGGGCTTTTGGCATGGAGAA 2820  
2852 AAAAAGATATTCGAAGATCTTCTCGAGGAAGAGAGTCAATGAAGACACAAATGGGATA 2911  
2821 AAAAAGATATTCGAAGATCTTCTCGAGGAAGAGAGTCAATGAAGACACAAATGGGATA 2880  
2912 CTTCACTGATAGCAAAAATCTCGGAGGCAACTAAAGATACATTTGCAGATTCCTCAG 2971  
2881 CTTCACTGATAGCAAAAATCTCGGAGGCAACTAAAGATACATTTGCAGATTCCTCAG 2940  
2972 ATATGTAATAAATCTTAATAGCAAGTTTGGATTTCATCGCGGAAAGTCCCTGCTCA 3031  
2941 ATATGTAATAAATCTTAATAGCAAGTTTGGATTTCATCGCGGAAAGTCCCTGCTCA 3000  
3032 CATGCTCAGATGATGACCGGATTTGTTATGCAAGAACTGCAAGATATGTTCCCTGAAGA 3091  
3001 CATGCTCAGATGATGACCGGATTTGTTATGCAAGAACTGCAAGATATGTTCCCTGAAGA 3060  
3092 ATTTGCAAGACGTCATTTCAAAAGTGGCCATTTCTGAGGATATGCAAGTTTGGCTTCTC 3151  
3061 ATTTGCAAGACGTCATTTCAAAAGTGGCCATTTCTGAGGATATGCAAGTTTGGCTTCTC 3120  
3152 TTATTTTATTTATCTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTTTGATGA 3211  
3121 TTATTTTATTTATCTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTTTGATGA 3180  
3212 AGTTGATACAGATCAATCTGGTGTCTTCTGACAGAGAAATCCGAACACTGGCTACCAG 3271  
3181 AGTTGATACAGATCAATCTGGTGTCTTCTGACAGAGAAATCCGAACACTGGCTACCAG 3240  
3272 AATTACAGAACTGGGTTAAGTTTGCAGGATTTGACAGGTCTGGAACACATGCTAATAA 3331  
3241 AATTACAGAACTGGGTTAAGTTTGCAGGATTTGACAGGTCTGGAACACATGCTAATAA 3300  
3332 TTGCTCAAAATGCTTCTGCTGTATATCACGACGATAAATAATTCACCAACTCAGGA 3391  
3301 TTGCTCAAAATGCTTCTGCTGTATATCACGACGATAAATAATTCACCAACTCAGGA 3360  
3392 ATCTTACTATGATCCCAACCTGCCACCGGTCACTTAAAGTCTAGTAAACAACTGTAAAC 3451  
3361 ATCTTACTATGATCCCAACCTGCCACCGGTCACTTAAAGTCTAGTAAACAACTGTAAAC 3420  
3452 AGTAACTACAAAATCCCAAGACATATAGGACAAAACAAATATAGTTTGAATCAT 3511  
3421 AGTAACTACAAAATCCCAAGACATATAGGACAAAACAAATATAGTTTGAATCAT 3480  
3512 GGGAGAGAGAAATCGTTTTAAATGATTCGTACCAGCTTTCTCATGCTGGTTGGCCCA 3571  
3481 GGGAGAGAGAAATCGTTTTAAATGATTCGTACCAGCTTTCTCATGCTGGTTGGCCCA 3540  
3572 GTTGGATGACATAGAAAACCTAGGAAGTTTGGTCTGATGATGACAACTGACCA 3631  
3541 GTTGGATGACATAGAAAACCTAGGAAGTTTGGTCTGATGATGACAACTGACCA 3600  
3632 CAATCATAAAGATGCTCAGACAGTGAAGGCTTGTCTCAGGGACTTCTATGAATCCATGTT 3691  
3601 CAATCATAAAGATGCTCAGACAGTGAAGGCTTGTCTCAGGGACTTCTATGAATCCATGTT 3660  
3692 CCCATACCTTCCCAATTTGAACTGCCAGAGAGATTCGAAACCGTTTCTTATATGCA 3751  
3661 CCCATACCTTCCCAATTTGAACTGCCAGAGAGATTCGAAACCGTTTCTTATATGCA 3720  
3752 TGAGCTGAGGAATGGAGGGCTTATTCGAGACAAATTTGAAGTTTGGACCCCATTTGTACT 3811  
3721 TGAGCTGAGGAATGGAGGGCTTATTCGAGACAAATTTGAAGTTTGGACCCCATTTGTACT 3780  
3812 AGCAACATGATATGTTTACTATATCTCATTTTCTGAGAGAGTAAATTCGACTTAA 3871

3781 AGCAACATTCATTATGTTTACTATATTTCTCATTTTCTGCTGAGCAGTTAATTCGACTTAA 3840  
3872 GCGGAGATATTTCCAGAGGAGGATACAAAAGACTAGTCCCAATCGAATCAGAGT 3931  
3841 GCGGAGATATTTCCAGAGGAGGATACAAAAGACTAGTCCCAATCGAATCAGAGT 3900  
3932 ATAGAGATCTTCATTGAAAACCATCTACCTCAGCATTACTGAGCAATTTTAAACTCA 3991  
3901 ATAGAGATCTTCATTGAAAACCATCTACCTCAGCATTACTGAGCAATTTTAAACTCA 3960  
3992 GCTTCACAGAGATGCTTTTGTGATGTGCTTTAGCAGTTTGGCCCGAAGAGGAAATA 4051  
3961 GCTTCACAGAGATGCTTTTGTGATGTGCTTTAGCAGTTTGGCCCGAAGAGGAAATA 4020  
4052 TCAGTACCATGCTGTTTGGCATGAATATAGCCCACTGACTAGGAATTTTAACTCA 4111  
4021 TCAGTACCATGCTGTTTGGCATGAATATAGCCCACTGACTAGGAATTTTAACTCA 4080  
4112 ACCCACTGAAAATTTGTGTCGAGCAGCTCTGAACTGATTTTACTTTTAAAGAAATTTGC 4171  
4081 ACCCACTGAAAATTTGTGTCGAGCAGCTCTGAACTGATTTTACTTTTAAAGAAATTTGC 4140  
4172 TCATGAGCTGTCCTCTTTTATATAAAGGCTCTACTGACAGAGACAGCTGTTAAATTC 4231  
4141 TCATGAGCTGTCCTCTTTTATATAAAGGCTCTACTGACAGAGACAGCTGTTAAATTC 4200  
4232 CCACAGCAATCAATTCGAGCACTAACTTTATTAGGAGAGCCCTATGCGAGCTGGAGTATT 4291  
4201 CCACAGCAATCAATTCGAGCACTAACTTTATTAGGAGAGCCCTATGCGAGCTGGAGTATT 4260  
4292 GCTAAGAGCTCCAGCTCTTTGCTATCCAAAGCCTTTGCTAAAGTTTGGCACTTTT- 4350  
4261 GCTAAGAGCTCCAGCTCTTTGCTATCCAAAGCCTTTGCTAAAGTTTGGCACTTTTCT 4320  
4351 TTTTTCATTTCCCATTTTAAAGTAGTTA-CTAAGTTTAACTAGTTATT 4395  
4321 TTTTTCATTTCCCATTTTCAAGTAGTTTAACTAGTTTAACTAGTTATT 4366

RESULT 6  
AX468106 4277 bp DNA linear PAT 16-JUL-2002  
Sequence 29 from Patent WO0250279.  
LOCUS AX468106  
DEFINITION AX468106  
ACCESSION AX468106.1 GI:21900980  
VERSION  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1  
Baughin,M.R., Lu,Y., Arvizu,C., Ramkumar,J., Yao,M.G.,  
Policky,J.L., Walla,N.K., Tribouley,K.M., Yue,H., Batra,S.,  
Ding,L., Lal,P.G., Borowsky,M.L., Lu,D.A., Gandhi,A.R.,  
Griffin,J.A., Xu,Y., Azimzai,Y., Gietzen,K.J., Tang,Y.T.,  
Warren,B.A., Mason,P.M., Burford,N., Hafalia,A.J., Lee,E.A.,  
Yang,J., Gorvad,A.E., Emerling,B.M., Marquis,J.P., Lee,S.Y.,  
Swarnakar,A. and Reddy,R.  
Nucleic acid-associated proteins  
Patent: WO 0250279-A 29 27-JUN-2002;  
Incyte Genomics, Inc. (US)  
Incyte Genomics, Inc. (US)  
Location/Qualifiers  
1. 4277  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: 7503560CB1"

TITLE  
JOURNAL  
FEATURES  
Source

ORIGIN  
Query Match 72.5%; Score 4057; DB 6; Length 4277;  
Best Local Similarity 95.7%; Pred. No. 0;  
Matches 4264; Conservative 0; Mismatches 11; Indels 181; Gaps 2;



26	GGAGCTGCAATGAGCGGCGCCCGAGAGGCTGTGACTTGC	CGCGCGCGCGCCGACCGGGGCC	85
2	CGAGCTGCAATGAGCGGCGCCCGAGAGGCTGTGACTTGC	CGCGCGCGCGCCGACCGGGGCC	61
86	CTGTAATGCGGCTCGCTCAGAGCGCGCGCGCGCGGCTC	CAGCTCCTCGGGGCGTG	145
62	CTGTAATGCGGCTCGCTCAGAGCGCGCGCGCGCGGCTC	CAGCTCCTCGGGGCGTG	121
146	GGTGGCGGTGAAGGGTGATGCTGTTCAAGTCTCTGCAG	AGACAAAACCTATACCTGGCT	205
122	GGTGGCGGTGAAGGGTGATGCTGTTCAAGTCTCTGCAG	AGACAAAACCTATACCTGGCT	181
206	GTCCACACAGTATGGGCTCCTAGGTGCTTCTTGGGCTG	CTTGTGTACCATCGTCTCCGC	265
182	GTCCACACAGTATGGGCTCCTAGGTGCTTCTTGGGCTG	CTTGTGTACCATCGTCTCCGC	241
266	CTTCAGTTTCGGAGAGGGTGGTCTCGAAATGAGACCGAG	ATCAATACCATTGTTTGTGA	325
242	CTTCAGTTTCGGAGAGGGTGGTCTCGAAATGAGACCGAG	ATCAATACCATTGTTTGTGA	301
326	TTCTCTATAGACAAATATTGCTTGGAAAGTCCCTTTCA	GAATCGGCTTTGTCTGCCATGCC	385
302	TTCTCTATAGACAAATATTGCTTGGAAAGTCCCTTTCA	GAATCGGCTTTGTCTGCCATGCC	361
386	GATTGACGTTGTTTACACCTGGTGAATGGCACAGATCTT	GAACTACTGAAGGAACCTACA	445
362	GATTGACGTTGTTTACACCTGGTGAATGGCACAGATCTT	GAACTACTGAAGGAACCTACA	421
446	GCAGTTCAGAGAACAGATGAGAGAGGAGCAGAAAGCA	ATGAGAGAAATCCTTGGGAAAA	505
422	GCAGTTCAGAGAACAGATGAGAGAGGAGCAGAAAGCA	ATGAGAGAAATCCTTGGGAAAA	481
506	CACACGGAACTACTTAAGAAGAGTGAGAAAGCAGTTA	GAGTGTGCTAACACACTGCAT	565
482	CACACGGAACTACTTAAGAAGAGTGAGAAAGCAGTTA	GAGTGTGCTAACACACTGCAT	541
566	TAAGTGCCTTACCAATGCTCTCAGTTGTTACTGGACCC	AGCCCTGCCAGCAACATCACCTGAAAGAGT	625
542	TAAGTGCCTTACCAATGCTCTCAGTTGTTACTGGACCC	AGCCCTGCCAGCAACATCACCTGAAAGAGT	601
626	GCATCTCTTTATCCTTCTTTTCACTTCCGACGTGAC	ATTTTCAATGTTGCAAAACCAAA	685
602	GCATCTCTTTATCCTTCTTTTCACTTCCGACGTGAC	ATTTTCAATGTTGCAAAACCAAA	661
686	AAACCTCTTACCAATGCTCTCAGTTGTTGTTTTTGA	CAGTACTTAAGGATGTTGAAGATGC	745
662	AAACCTCTTACCAATGCTCTCAGTTGTTGTTTTTGA	CAGTACTTAAGGATGTTGAAGATGC	721
746	CCACTCTGACACTGCTTTAAAGGAAATAGCAGACAGAC	AGTATGGAGGGGTACTTGCACAC	805
722	CCACTCTGACACTGCTTTAAAGGAAATAGCAGACAGAC	AGTATGGAGGGGTACTTGCACAC	781
806	AGATAAAGAAGTCCCTGGATTAAGTGTCTAATGCAAG	ATTTGGCTTTTCTCAGTGGATTTCC	865
782	AGATAAAGAAGTCCCTGGATTAAGTGTCTAATGCAAG	ATTTGGCTTTTCTCAGTGGATTTCC	841
866	ACCAACATTCAGAGGAAACAAATCAACTAAAAAACAA	AATTTGCCAGAAAAATCTTTCTCTAA	925
842	ACCAACATTCAGAGGAAACAAATCAACTAAAAAACAA	AATTTGCCAGAAAAATCTTTCTCTAA	901
926	AGTCAAACTGTTGCAGTCTGTTATTCAGAGGCCAGTG	TAGCGCTTCTTAAAACTCAATAACC	985
902	AGTCAAACTGTTGCAGTCTGTTATTCAGAGGCCAGTG	TAGCGCTTCTTAAAACTCAATAACC	961
986	CAAGGATTTTCAAGAAATGAATAAGCAAACTAAGAAG	AACATGACCATTTGATGAAAAAGA	1045
962	CAAGGATTTTCAAGAAATGAATAAGCAAACTAAGAAG	AACATGATGAAAAAGA	1021
1046	ACTGACCATTAAGTCTGCATATTTATTTATTTGGGAT	CTTGAGGCCATCAGCCAGTCTAAGCA	1105
1022	ACTGACCATTAAGTCTGCATATTTATTTATTTGGGAT	CTTGAGGCCATCAGCCAGTCTAAGCA	1081
1106	GGATGAAGACATCTCTCGCCAGCTGTTTTTGAAGATAC	GGAAGAACTGAGGTACTCATTTGG	1165



QY 4406 AGTATAACGAATTGGGATGCTCTAAR-CCTATTTTATAGATGCTTATTTAAATATATGACG 4464  
Db 4202 AGTATAACGAATTGGGATGCTCTAARCCCTATTTTATAGATGCTTATTTAAATATATGACG 4261  
QY 4465 AATATCACCTCTTAT 4480  
Db 4262 ATTTCACCTCTTTT 4277

RESULT 7  
LOCUS AR300471 3783 bp DNA linear PAT 12-JUN-2003  
DEFINITION Sequence 20 from patent US 6537785.  
ACCESSION AR300471  
VERSION AR300471.1 GI:31687913  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 3783)  
AUTHORS Canfield, W.M.  
TITLE Methods of treating lysosomal storage diseases  
JOURNAL Patent: US 6537785-A 20 25-MAR-2003;  
FEATURES  
location/Qualifiers  
1..3783  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Query Match 67.2%; Score 3761; DB 6; Length 3783;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 3767; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
QY 165 ATGCTGTTCAAGCTCCTGAGAGACAAACCTATACCTGCTCCACAGATGATGGGCTC 224  
Db 7 ATGGGGTTCAAGCTCTTGAGAGACAAACCTATACCTGCTCCACAGATGATGGGCTC 66  
QY 225 TACGTGTGCTTTGGGGCTGCTGTACCAATGCTCGGCTCCAGTTCGGAGAGTG 284  
Db 67 TACGTGTGCTTTGGGGCTGCTGTACCAATGCTCGGCTCCAGTTCGGAGAGTG 126  
QY 285 GTTCTGGATGAGCGGAGATCAATACCATGTTTGTGTTGATCTCTATAGACATATT 344  
Db 127 GTTCTGGATGAGCGGAGATCAATACCATGTTTGTGTTGATCTCTATAGACATATT 186  
QY 345 GCTGGAAGTCTTTTCAAGTCCGCTTTGCTGCCATGCGGATGAGCTGTTTACACC 404  
Db 187 GCTGGAAGTCTTTTCAAGTCCGCTTTGCTGCCATGCGGATGAGCTGTTTACACC 246  
QY 405 TGGGTGAATGGACAGATCTTGAATCTAGAGGAATCTACAGAGGTACAGAGACAGATG 464  
Db 247 TGGGTGAATGGACAGATCTTGAATCTAGAGGAATCTACAGAGGTACAGAGACAGATG 306  
QY 465 GAGGAGGACAGAAAGCAATGAGAGAAATCCCTTGGGAAAAACACAAACGAACTTACTAAG 524  
Db 307 GAGGAGGACAGAAAGCAATGAGAGAAATCCCTTGGGAAAAACACAAACGAACTTACTAAG 366  
QY 525 AAGAGTGAGACAGATGATGATGTTGCTTAACACACTGCAATTAAGGTCCCAATGCTTGT 584  
Db 367 AAGAGTGAGACAGATGATGATGTTGCTTAACACACTGCAATTAAGGTCCCAATGCTTGT 426  
QY 585 CTGGACCCAGCCCTGCGAGCAACATCAACCTGAGAGGACGTGCCATCTCTTATCCTTCT 644  
Db 427 CTGGACCCAGCCCTGCGAGCAACATCAACCTGAGAGGACGTGCCATCTCTTATCCTTCT 486  
QY 645 TTTTCATCTGCCAGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGTC 704  
Db 487 TTTTCATCTGCCAGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGTC 546  
QY 705 TCAGTTGTTGTTTGGACAGTACTAAGGATGTTGAAGTGGCCACTCTGGACTGTTTAA 764  
Db 547 TCAGTTGTTGTTTGGACAGTACTAAGGATGTTGAAGTGGCCACTCTGGACTGTTTAA 606

QY 765 GGAATAGCAGACAGACAGTATGAGGGGTACTTGCACACAGATATAAGAGTCCCTGGA 824  
Db 607 GGAATAGCAGACAGACAGTATGAGGGGTACTTGCACACAGTAAAGAGTCCCTGGA 666  
QY 825 TTAGTGCTAATGCAAGATTTGGCTTTCTAGTGGATTTCCACCAACATTTCAAGGAAACA 884  
Db 667 TTAGTGCTAATGCAAGATTTGGCTTTCTAGTGGATTTCCACCAACATTTCAAGGAAACA 726  
QY 885 AATCACTAAACAAATTTGCCAGAAATCTTTCTTAAAGTCAAACTGTTGCAAGTTG 944  
Db 727 AATCACTAAACAAATTTGCCAGAAATCTTTCTTAAAGTCAAACTGTTGCAAGTTG 786  
QY 945 TATTTCAGAGCCAGTGTAGCGCTTTCTAAACCTGAATAACCCCAAGGATTTTCAAGAAATTG 1004  
Db 787 TATTTCAGAGCCAGTGTAGCGCTTTCTAAACCTGAATAACCCCAAGGATTTTCAAGAAATTG 846  
QY 1005 AATAAGCAAACTAAGAAAGACATGACATGATGAGAAAGAACTGACCATTAAGTCTGCA 1064  
Db 847 AATAAGCAAACTAAGAAAGACATGACATGATGAGAAAGAACTGACCATTAAGTCTGCA 906  
QY 1065 TATTATTATGGGATCTGAGCGCATCAGCCAGTCTAAGCAGAGATGAAGACATCTCTGCC 1124  
Db 907 TATTATTATGGGATCTGAGCGCATCAGCCAGTCTAAGCAGAGATGAAGACATCTCTGCC 966  
QY 1125 AGTCGTTTGAAGATAACGAAGAACTGAGGTAATCTATTTGGGATCTATTCGAGAGGCAATGCA 1184  
Db 967 AGTCGTTTGAAGATAACGAAGAACTGAGGTAATCTATTTGGGATCTATTCGAGAGGCAATGCA 1026  
QY 1185 CCATGGTTCGGAATATTTTCAATGTCACCAACGGCAGATTCATCTCTGGCTGAACCTT 1244  
Db 1027 CCATGGTTCGGAATATTTTCAATGTCACCAACGGCAGATTCATCTCTGGCTGAACCTT 1086  
QY 1245 GACAATCTCGAGTGAACAATAGTAACACACACAGGATGTTTTCGAAATTTGAGCCACTTG 1304  
Db 1087 GACAATCTCGAGTGAACAATAGTAACACACACAGGATGTTTTCGAAATTTGAGCCACTTG 1146  
QY 1305 CCTACCTTTAGTTCACCTGCTATTTGAAGTCAATTCATCGCATCGAAGGGCTGTCCCAG 1364  
Db 1147 CCTACCTTTAGTTCACCTGCTATTTGAAGTCAATTCATCGCATCGAAGGGCTGTCCCAG 1206  
QY 1365 AAGTTTATTTACCTAAATGATGATGTCATGTTTGGGAAGGATGTCCTGGCCAGATGATTT 1424  
Db 1207 AAGTTTATTTACCTAAATGATGATGTCATGTTTGGGAAGGATGTCCTGGCCAGATGATTT 1266  
QY 1425 TACAGTCACTCCAAAGCCAGAAAGGTTTATTTGACATGGCTGTGCAAACTGTGCCGAG 1484  
Db 1267 TACAGTCACTCCAAAGCCAGAAAGGTTTATTTGACATGGCTGTGCAAACTGTGCCGAG 1326  
QY 1485 GGCTGCCAGGTTTCCCTGGATTAAGGATGCTATTTGTGCAAGAGCTTGTAAATTTACAGCC 1544  
Db 1327 GGCTGCCAGGTTTCCCTGGATTAAGGATGCTATTTGTGCAAGAGCTTGTAAATTTACAGCC 1386  
QY 1545 TGCGATTTGGATGCTGGGATTCCTCTGGAACACAGTGGAGGAGTCCGTATATTGAGGA 1604  
Db 1387 TGCGATTTGGATGCTGGGATTCCTCTGGAACACAGTGGAGGAGTCCGTATATTGAGGA 1446  
QY 1605 GGTGGAGTACTGGGATTTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGAAATAAC 1664  
Db 1447 GGTGGAGTACTGGGATTTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGAAATAAC 1506  
QY 1665 AGTGTCTCTTACCTGATTAACAGGATGCGAAATTCCTGGCTCGCTGATAGTTCTGTGAC 1724  
Db 1507 AGTGTCTCTTACCTGATTAACAGGATGCGAAATTCCTGGCTCGCTGATAGTTCTGTGAC 1566  
QY 1725 CAAGCATGCAATGCTCTTGTCTGGGTTTGGTGGGACHTGGGGCAAGATCAATTT 1784  
Db 1567 CAAGCATGCAATGCTCTTGTCTGGGTTTGGTGGGACHTGGGGCAAGATCAATTT 1626  
QY 1785 CATGAATTTGATAAAGTATCTTCTCCAAACACAGACTCACTATATTATTTCCAAAGGT 1844  
Db 1627 CATGAATTTGATAAAGTATCTTCTCCAAACACAGACTCACTATATTATTTCCAAAGGT 1686  
QY 1845 GAATGCTGCTTTATTTACGCTTTTGAGAGTGTAGCCAAAGAGAGGATTTGAGGTGCCTAT 1904

Db 1687 GAATCCCTGCGCTATTTTTCAGCTTTGCGAAGTAGCCAAAGAGGAGTTGAAGTGCCTAT 1746  
QY 1905 AGTGACAATCCAAATATTCGACATGCTTCTATTGCGCAACAGAGTGGAAGAACCATCCACCTC 1964  
Db 1747 AGTGACAATCCAAATATTCGACATGCTTCTATTGCGCAACAGAGTGGAAGAACCATCCACCTC 1806  
QY 1965 ATAATGCGACAGTGGAATGAATGCGCACACATATATTTTAACTCTCAGCTTTCAAAATACA 2024  
Db 1807 ATAATGCGACAGTGGAATGAATGCGCACACATATATTTTAACTCTCAGCTTTCAAAATACA 1866  
QY 2025 AACGATGAAGAGTTCAAAATGCGAGATAACAGTGGAGGTGGACACAAAGGGAGGACCAAAA 2084  
Db 1867 AACGATGAAGAGTTCAAAATGCGAGATAACAGTGGAGGTGGACACAAAGGGAGGACCAAAA 1926  
QY 2085 CTGAATCTCAGCGCCAGAGGGTTACGAAATTTAGTTAGTCCCATACACTTCTTCCA 2144  
Db 1927 CTGAATCTCAGCGCCAGAGGGTTACGAAATTTAGTTAGTCCCATACACTTCTTCCA 1986  
QY 2145 GAGCGGAAATCCTTTTTCAGGATATCCCAAGAGAAAGCGTTCCCGAAGTTTAAAGAGA 2204  
Db 1987 GAGCGGAAATCCTTTTTCAGGATATCCCAAGAGAAAGCGTTCCCGAAGTTTAAAGAGA 2046  
QY 2205 CATGATGTTAACTCAACAGGAGAGCCCGAGGAGAGGTGAAATTTCCCTGGTAAATATT 2264  
Db 2047 CATGATGTTAACTCAACAGGAGAGCCCGAGGAGAGGTGAAATTTCCCTGGTAAATATT 2106  
QY 2265 TCACCTCTCCAAAGAGCGCCAGTGTGCTCAATACCTTGGATTTGCACTGGACAT 2324  
Db 2107 TCACCTCTCCAAAGAGCGCCAGTGTGCTCAATACCTTGGATTTGCACTGGACAT 2166  
QY 2325 GGAGACATCACTTTGAAAGGATACAAATTTGTCGAAGTCCAGCTTGCCTGAGATCATTTCTG 2384  
Db 2167 GGAGACATCACTTTGAAAGGATACAAATTTGTCGAAGTCCAGCTTGCCTGAGATCATTTCTG 2226  
QY 2385 ATGAATCTCAGAGCATGCTAAATATAAATAATCAAGCTATATAACAGATGAAACAAATGAC 2444  
Db 2227 ATGAATCTCAGAGCATGCTAAATATAAATAATCAAGCTATATAACAGATGAAACAAATGAC 2286  
QY 2445 AGTTTGGTGGCTCCACAGGAAAGACAGGTTTCAATAAGCATCTTGCACAAACAGCTTAGGA 2504  
Db 2287 AGTTTGGTGGCTCCACAGGAAAGACAGGTTTCAATAAGCATCTTGCACAAACAGCTTAGGA 2346  
QY 2505 GTGCTGAAAGATGCGAGAGTTGACTTTTCTCGAGTGGAGTGAAGTGAATGGTCAAT 2564  
Db 2347 GTGCTGAAAGATGCGAGAGTTGACTTTTCTCGAGTGGAGTGAAGTGAATGGTCAAT 2406  
QY 2565 GACCAAGGCTCAGATCCACCCCTGAGCTTGGAGACCCAGCAGCAAGATTTAGAGTGGAACT 2624  
Db 2407 GACCAAGGCTCAGATCCACCCCTGAGCTTGGAGACCCAGCAGCAAGATTTAGAGTGGAACT 2466  
QY 2625 CACACCCCAAAACCATAGGCGGAAATGTGACAAAGAAAGAGCCCCCATCTCTGATTTGTT 2684  
Db 2467 CACACCCCAAAACCATAGGCGGAAATGTGACAAAGAAAGAGCCCCCATCTCTGATTTGTT 2526  
QY 2685 CCACTGGAGGCCAGATGACAAAGAAAGAAATCAAGGAAAGAAAGAAAGAACAGT 2744  
Db 2527 CCACTGGAGGCCAGATGACAAAGAAAGAAATCAAGGAAAGAAAGAAAGAACAGT 2586  
QY 2745 AGAATGGAGGAAATGCTGAAATACATAGGCGCTTACTGAAAGTGTACTTGGAGAAAG 2804  
Db 2587 AGAATGGAGGAAATGCTGAAATACATAGGCGCTTACTGAAAGTGTACTTGGAGAAAG 2646  
QY 2805 CTGACGATTTACAGATAGTTACTTGGGCTTTTGGCATGGGAGAAAGAAAGTATTTTC 2864  
Db 2647 CTGACGATTTACAGATAGTTACTTGGGCTTTTGGCATGGGAGAAAGAAAGTATTTTC 2706  
QY 2865 CAAATCTTCTCGAGGAGAGAGTCAATTTGAAGACAAATTTGGCATCTTCACTGTATGC 2924  
Db 2707 CTAGATCTTCTCGAGGAGAGAGTCAATTTGAAGACAAATTTGGCATCTTCACTGTATGC 2766  
QY 2925 AAAAATCTGGAGGCAACTAAAGATACATTTGCGAGATCCCTTCAGATATGTAATAAA 2984

Db 2767 AAGNATCTGGGAGGCACTAAAGATACATTTGCGAGATCCCTCAGATATGTAATAAA 2826  
QY 2985 ATTCTAAATAGCAAGTTTGGATTTCATCGCGGAAAGTCCCTGCTCAATGCTCTCACATG 3044  
Db 2827 ATTCTAAATAGCAAGTTTGGATTTCATCGCGGAAAGTCCCTGCTCAATGCTCTCACATG 2886  
QY 3045 ATTGACCGGATTTGTTATGCAAGAACTGCAAGATATGTTCCCTGGAAGATTTGACAAGACG 3104  
Db 2887 ATTGACCGGATTTGTTATGCAAGAACTGCAAGATATGTTCCCTGGAAGATTTGACAAGACG 2946  
QY 3105 TCATTTCAAAAGTGGCCATTCTCAGGATATGCAAGTTTGGCTTCTCTTATTTTATTTAT 3164  
Db 2947 TCATTTCAAAAGTGGCCATTCTCAGGATATGCAAGTTTGGCTTCTCTTATTTTATTTAT 3006  
QY 3165 CTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGTATACAGAT 3224  
Db 3007 CTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGTATACAGAT 3066  
QY 3225 CAATCTGGTGTCTTGTCTGACAGAGAAATCCGAAACACTGGCTTACAGAAATTCAGGAACTG 3284  
Db 3067 CAATCTGGTGTCTTGTCTGACAGAGAAATCCGAAACACTGGCTTACAGAAATTCAGGAACTG 3126  
QY 3285 CCGTTAAGTTTGCAGGATTTGACAGTCTGGAACACATGCTTAATAAATGCTCAAAAATG 3344  
Db 3127 CCGTTAAGTTTGCAGGATTTGACAGTCTGGAACACATGCTTAATAAATGCTCAAAAATG 3186  
QY 3345 CTTCTGCTGATATCAGCGAGCTAAATAATATTCCACCAACTCAGGAAATTCCTACTATGAT 3404  
Db 3187 CTTCTGCTGATATCAGCGAGCTAAATAATATTCCACCAACTCAGGAAATTCCTACTATGAT 3246  
QY 3405 CCCAACTGCGCCACCGGTCACTAAAGTCTAGTAAACAAACTGTAAACCCAGTAACTGACAAA 3464  
Db 3247 CCCAACTGCGCCACCGGTCACTAAAGTCTAGTAAACAAACTGTAAACCCAGTAACTGACAAA 3306  
QY 3465 ATCCACAGAGCATATAAGGACAAACAAATATAGTTTGGAAATCATGGGAGAGAAAG 3524  
Db 3307 ATCCACAGAGCATATAAGGACAAACAAATATAGTTTGGAAATCATGGGAGAGAAAG 3366  
QY 3525 ATCGCTTTTAAATGATTCGTACCAACGTTTCTCATGTTGGTGGCCAGTTGGATGACATA 3584  
Db 3367 ATCGCTTTTAAATGATTCGTACCAACGTTTCTCATGTTGGTGGCCAGTTGGATGACATA 3426  
QY 3585 AGAAAAACCTTAGAAGTTTGGTTCCTGATGACCAACTGACCAACTCAATCAATAGAT 3644  
Db 3427 AGAAAAACCTTAGAAGTTTGGTTCCTGATGACCAACTGACCAACTCAATCAATAGAT 3486  
QY 3645 GCTCAGACAGTGAAGGCTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCC 3704  
Db 3487 GCTCAGACAGTGAAGGCTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCC 3546  
QY 3705 CAAATTTGAATCCCAAGAGAGTATCGAAAACGTTTCTTCTCATATGCAATGACCTGAGGAA 3764  
Db 3547 CAAATTTGAATCCCAAGAGAGTATCGAAAACGTTTCTTCTCATATGCAATGACCTGAGGAA 3606  
QY 3765 TGGAGGCTTATCGAGACAAATTTGAAGTTTGGACCCATTTGTACTAGCAACTTGAAT 3824  
Db 3607 TGGAGGCTTATCGAGACAAATTTGAAGTTTGGACCCATTTGTACTAGCAACTTGAAT 3666  
QY 3825 ATGTTTACTATATCTCATTTTCTGCTGAGCAGTTTAAATGACCTTTAAACGGAAGATATTT 3884  
Db 3667 ATGTTTACTATATCTCATTTTCTGCTGAGCAGTTTAAATGACCTTTAAACGGAAGATATTT 3726  
QY 3885 CCCAAGAGGAGTACACAAAGAGCTAGTCCCATCGAATCAGAGTATAGAGATC 3941  
Db 3727 CCCAAGAGGAGTACACAAAGAGCTAGTCCCATCGAATCAGAGTATAGAGATC 3783

RESULT 8

AR428748

LOCUS

DEFINITION

ACCESSION

AR428748

Sequence 20 from patent US 6642038.

AR428748

3783 bp

DNA

linear

PAT 18-DEC-2003

## KEYWORDS

SOURCE  
ORGANISMUnknown.  
Unknown.

REFERENCE

1 (bases 1 to 3783)

AUTHORS

Canfield, W.M.

TITLE

GlucNac phosphotransferase of the lysosomal targeting pathway

JOURNAL

Patent: US 6642039-A 20 04-NOV-2003;

FEATURES

Location/Qualifiers

1..3783

/organism="unknown"

/mol\_type="genomic DNA"

## ORIGIN

Query Match 67.2%; Score 3761; DB 6; Length 3783;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 3767; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 165 ATGCTGTTCAAGCTCCTCGAGAGACAAACCTATACCTCGCTGTCCACAGGTATGGGCTC 224  
DB 7 ATGGGTTCAAGCTCCTCGAGAGACAAACCTATACCTCGCTGTCCACAGGTATGGGCTC 66

QY 225 TACGTGTGCTTCTGGGCGTGGTGTCAACATCGTCTCCGCTCCGCTTCCAGTTCGGAGAGTG 284  
DB 67 TACGTGTGCTTCTGGGCGTGGTGTCAACATCGTCTCCGCTCCGCTTCCAGTTCGGAGAGTG 126

QY 285 GTTCTGGAATGAGCGGAGATCAATACCATGTTTTGTTGTTGATTCCTATAGAGACAATATT 344  
DB 127 GTTCTGGAATGAGCGGAGATCAATACCATGTTTTGTTGTTGATTCCTATAGAGACAATATT 186

QY 345 GCTGAAAGTCCCTTTTCAAGATCGGCTTTGCTGCCATGCCGATGCGGATTTGACCTTACACC 404  
DB 187 GCTGAAAGTCCCTTTTCAAGATCGGCTTTGCTGCCATGCCGATTTGACCTTACACC 246

QY 405 TGGGTGAATGGACAGATCTTGAATCTAGAGNACTACAGCAGGTACAGAGACAGATG 464  
DB 247 TGGGTGAATGGACAGATCTTGAATCTAGAGNACTACAGCAGGTACAGAGACAGATG 306

QY 465 GAGGAGGAGCGAAGCAATGAGAGAAATCCTTGGGAAAAACAAACGACCTTACTAAG 524  
DB 307 GAGGAGGAGCGAAGCAATGAGAGAAATCCTTGGGAAAAACAAACGACCTTACTAAG 366

QY 525 AAGATGAGACAGTCTAGAGTGTGTTGTAACACATGCATTAAGGTGCCAATGCTTGTGA 584  
DB 367 AAGATGAGACAGTCTAGAGTGTGTTGTAACACATGCATTAAGGTGCCAATGCTTGTGC 426

QY 585 CTGGACCCAGCCCTGCCAGCAACATCACCTGAGGACGTGCCATCTCTTTATCCTTCT 644  
DB 427 CTGGACCCAGCCCTGCCAGCAACATCACCTGAGGACGTGCCATCTCTTTATCCTTCT 486

QY 645 TTTCATCTGCCAGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTTACCAATGTC 704  
DB 487 TTTCATCTGCCAGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTTACCAATGTC 546

QY 705 TCAGTTGTGTTTTGACAGTACTAAGGATGTTGAGATGCCACTCTGGACTGCTTAAA 764  
DB 547 TCAGTTGTGTTTTGACAGTACTAAGGATGTTGAGATGCCACTCTGGACTGCTTAAA 606

QY 765 GGAATAGCAGACACAGTATGAGGGGGTACTTGACACAGATAAAGAGTCCCTGGA 824  
DB 607 GGAATAGCAGACACAGTATGAGGGGGTACTTGACACAGATAAAGAGTCCCTGGA 666

QY 825 TTAGTGCTAATGCAAGATTTGGCTTTCTGAGTGAATTTCCACCAACATTCGAAGAAACA 884  
DB 667 TTAGTGCTAATGCAAGATTTGGCTTTCTGAGTGAATTTCCACCAACATTCGAAGAAACA 726

QY 885 AATCAACTAAAAACAAAAATGCGAGAAAATCTTCTCTAAAGTCAAACTGTCAGTTG 944  
DB 727 AATCAACTAAAAACAAAAATGCGAGAAAATCTTCTCTAAAGTCAAACTGTCAGTTG 786

QY 945 TATTCAAGGGCAGTGTAGCGCTTCTAAAACCTGAATAACCCCAAGGATTTTCAAGAAATTG 1004  
DB 787 TATTCAAGGGCAGTGTAGCGCTTCTAAAACCTGAATAACCCCAAGGATTTTCAAGAAATTG 846

QY 1005 AATAAGCAAACTAAGAAAGAACATGACCAATGATGGAAGAAAGAACTGACCAATAGTCTCTGCA 1064  
DB 847 AATAAGCAAACTAAGAAAGAACATGACCAATGATGGAAGAAAGAACTGACCAATAGTCTCTGCA 906

QY 1065 TATTTATTTATGGATCTGAGCGCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCC 1124  
DB 907 TATTTATTTATGGATCTGAGCGCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCC 966

QY 1125 AGTCGTTTTGAAGATAACGAAGAACTGAGTACTCATTCGGATCTATCGAGAGGCAATGCA 1184  
DB 967 AGTCGTTTTGAAGATAACGAAGAACTGAGTACTCATTCGGATCTATCGAGAGGCAATGCA 1026

QY 1185 CCATGGTTCCGGAATATTTTTCATTTGTCACCAAGGGCAGATTCATCTCCCTGGTGAACCTT 1244  
DB 1027 CCATGGTTCCGGAATATTTTTCATTTGTCACCAAGGGCAGATTCATCTCCCTGGTGAACCTT 1086

QY 1245 GACAATCCTCGAGTGAACAATAGTAACACACACAGGATGTTTTCGAAATTTGAGCCACTTG 1304  
DB 1087 GACAATCCTCGAGTGAACAATAGTAACACACACAGGATGTTTTCGAAATTTGAGCCACTTG 1146

QY 1305 CCTACCTTTAGTTCACTGCTATTTGAAAGTCACTTCATCGCATCGAAGGGCTGTCCAG 1364  
DB 1147 CCTACCTTTAGTTCACTGCTATTTGAAAGTCACTTCATCGCATCGAAGGGCTGTCCAG 1206

QY 1365 AAGTTTATTTACCTAAATGATGATGTCATGTTTGGGAAGGATGTCCTGGCCAGATGATTTT 1424  
DB 1207 AAGTTTATTTACCTAAATGATGATGTCATGTTTGGGAAGGATGTCCTGGCCAGATGATTTT 1266

QY 1425 TACAGTCACTCCAAAGCCAGAGGTTTATTTGACATGCGCTGTGCAAACTGTGCGGAG 1484  
DB 1267 TACAGTCACTCCAAAGCCAGAGGTTTATTTGACATGCGCTGTGCAAACTGTGCGGAG 1326

QY 1485 GGTGCCCCAGGTTCTCGGATTAAGGATGCTATTTGCAAGGCTTGTATTAATTCAGCC 1544  
DB 1327 GGTGCCCCAGGTTCTCGGATTAAGGATGCTATTTGCAAGGCTTGTATTAATTCAGCC 1386

QY 1545 TGCATTTGGGATGCTGGGATTTGCTCTGAAACAGTGGAGGAGTGCCTATATTGCAAGA 1604  
DB 1387 TGCATTTGGGATGCTGGGATTTGCTCTGAAACAGTGGAGGAGTGCCTATATTGCAAGA 1446

QY 1605 GGTGAGGATCTGGGAGTATTTGAGTTGGACACCCCTGCGAGTTTGGTGGAGGAATAAAC 1664  
DB 1447 GGTGAGGATCTGGGAGTATTTGAGTTGGACACCCCTGCGAGTTTGGTGGAGGAATAAAC 1506

QY 1665 AGTGTCTCTTACTGTAATCAGGATGTCGGAATTCCTGCTCGCTGATTAAGTTCTGTGAC 1724  
DB 1507 AGTGTCTCTTACTGTAATCAGGATGTCGGAATTCCTGCTCGCTGATTAAGTTCTGTGAC 1566

QY 1725 CAAGCATGCAATGCTTTGCTGCTGGGTTTGAATGCTGGCGACTGTGGGCAAGATCAATTTT 1784  
DB 1567 CAAGCATGCAATGCTTTGCTGCTGGGTTTGAATGCTGGCGACTGTGGGCAAGATCAATTTT 1626

QY 1785 CATGAATTTGTATAAAGTGAATCCTTCTCCCAACCCAGACTCACTATATATTATCCAAAGGT 1844  
DB 1627 CATGAATTTGTATAAAGTGAATCCTTCTCCCAACCCAGACTCACTATATATTATCCAAAGGT 1686

QY 1845 GAATGCTGCTTTATTTTCAAGTTTGCAAGTGAAGTAGCCAAAGAGGAGTTGAAGTGCTAT 1904  
DB 1687 GAATGCTGCTTTATTTTCAAGTTTGCAAGTGAAGTAGCCAAAGAGGAGTTGAAGTGCTAT 1746

QY 1905 AGTGACATTCATTAATTCGATGCTTCTATTTGCCAACCAAGTGAAGAACCTCCACCTC 1964  
DB 1747 AGTGACATTCATTAATTCGATGCTTCTATTTGCCAACCAAGTGAAGAACCTCCACCTC 1806

QY 1965 ATAATGACAGTGGAAATGAATGCCACCAATATCAATTTTAAATCTCACTGTTTCAAAATACA 2024  
DB 1807 ATAATGACAGTGGAAATGAATGCCACCAATATCAATTTTAAATCTCACTGTTTCAAAATACA 1866

QY 2025 AACGNTGAGAGTTTCAAAATGAGATAACAGTGGAGTGAACAAAGGAGGAGGACCAAAA 2084  
DB 1867 AACGNTGAGAGTTTCAAAATGAGATAACAGTGGAGTGAACAAAGGAGGAGGACCAAAA 1926







1

1990

1027	Db	CCATGGGTTTCGGNAATATTTTCATTTGTCCACCAACGGCGAGATCCCATCTCTGGCTGAACCTT	1081
1245	QY	GACAACTCTCGAGTGAACAATAGTAACACACCAAGGATGTTTTTCGAAATTTTGAGCCACTTGG	1304
1087	Db	GACAACTCTCGAGTGAACAATAGTAACACACCAAGGATGTTTTTCGAAATTTTGAGCCACTTGG	1146
1305	QY	CCTACCTTTAGTTTCACTCTCTATTTGAAAGTCAATTCATCGCATCGAAGGGCTGTCCCAAG	1364
1147	Db	CCTACCTTTAGTTTCACTCTCTATTTGAAAGTCAATTCATCGCATCGAAGGGCTGTCCCAAG	1206
1365	QY	AAGTTTATTTACCTAAATGATGATGTCATGTTTGGGAAGGATGTCTTGGCCAGATGATTTT	1424
1207	Db	AAGTTTATTTACCTAAATGATGATGTCATGTTTGGGAAGGATGTCTTGGCCAGATGATTTT	1266
1425	QY	TACAGTCACTCCAAAGGCCAGAAAGTTTATTTTGAATGCGCTTGTCCCAAACCTGTCCCGAG	1484
1267	Db	TACAGTCACTCCAAAGGCCAGAAAGTTTATTTTGAATGCGCTTGTCCCAAACCTGTCCCGAG	1326
1485	QY	GGTGTGCCAGGTTCTCTGGATTAAGGATGCGCTATTTGTGAACAAGCTTTGTAATTAATTCAGCC	1544
1327	Db	GGCTGCCAGGTTCTCTGGATTAAGGATGCGCTATTTGTGAACAAGCTTTGTAATTAATTCAGCC	1386
1545	QY	TGCGATTTGGGATTTGGTGGGGAATTGCTCTCTGGAACAGTGGAGGAGTGCCTATATTGAGGA	1604
1387	Db	TGCGATTTGGGATTTGGTGGGGAATTGCTCTCTGGAACAGTGGAGGAGTGCCTATATTGAGGA	1446
1605	QY	GGTGGAGGTACTGGGAGTATTGGAGTTTGACACCCCTGCGAGTTTGGTGGAGGAATAAAC	1664
1447	Db	GGTGGAGGTACTGGGAGTATTGGAGTTTGACACCCCTGCGAGTTTGGTGGAGGAATAAAC	1506
1665	QY	AGTGTCTCTTACTGTAATCAGGATGTGCGAATTCCTGTGCTCGCTGATTAAGTTCTGTGAC	1724
1507	Db	AGTGTCTCTTACTGTAATCAGGATGTGCGAATTCCTGTGCTCGCTGATTAAGTTCTGTGAC	1566
1725	QY	CAAGCATGAANTGCTCTTGTCTGTGGTTTGTATGCTGGCGACTGTGGGCAAGATCATTTT	1784
1567	Db	CAAGCATGAANTGCTCTTGTCTGTGGTTTGTATGCTGGCGACTGTGGGCAAGATCATTTT	1626
1785	QY	CATGAATTGTATAAAGTGTACTCTCTCCAAACACGACTCACTATATTATTCCAAAGGT	1844
1627	Db	CATGAATTGTATAAAGTGTACTCTCTCCAAACACGACTCACTATATTATTCCAAAGGT	1686
1845	QY	GAAATGCTCTGCTTATTTACGCTTTTGACAGTAAGTAGCCAAAGAGGAGTTGAAAGTGCTTAT	1904
1687	Db	GAAATGCTCTGCTTATTTACGCTTTTGACAGTAAGTAGCCAAAGAGGAGTTGAAAGTGCTTAT	1746
1905	QY	AGTGACAACTCAATTAATTCGACATGCTTCTATTTGCCAACAAAGTGGAAAAACCATCCACCTC	1964
1747	Db	AGTGACAACTCAATTAATTCGACATGCTTCTATTTGCCAACAAAGTGGAAAAACCATCCACCTC	1806
1965	QY	ATAATGCAACAGTGGGAATGAATGCCACCACAATACATTTTAAATCTCACGTTTCAAAATACA	2024
1807	Db	ATAATGCAACAGTGGGAATGAATGCCACCACAATACATTTTAAATCTCACGTTTCAAAATACA	1866
2025	QY	AACGATGAAGATTCAAAATGACAGATAACAGTGGAGGTGGACACAGGAGGAGGCCAAAA	2084
1867	Db	AACGATGAAGATTCAAAATGACAGATAACAGTGGAGGTGGACACAGGAGGAGGCCAAAA	1926
2085	QY	CTGAATTTCTACGGCCAGGAAGGGTTACGAAAAATTTAGTTAGTCCCAATAACATTTCTTCCA	2144
1927	Db	CTGAATTTCTACGGCCAGGAAGGGTTACGAAAAATTTAGTTAGTCCCAATAACATTTCTTCCA	1986
2145	QY	GAGGCGGAATCTCTTTTTCAGGATATTCCCAAGAAAAACGCTTCCCGAAGTTTAAGGA	2204
1987	Db	GAGGCGGAATCTCTTTTTCAGGATATTCCCAAGAAAAACGCTTCCCGAAGTTTAAGGA	2046
2205	QY	CATGATGTTAACTCAACAAAGGAGAGCCACAGGAAGGTCGAAAAATTTCCCTCGTAAATATT	2264
2047	Db	CATGATGTTAACTCAACAAAGGAGAGCCACAGGAAGGTCGAAAAATTTCCCTCGTAAATATT	2106
2265	QY	TCACTCTCTTCCAAAAGAGCCGCAATGAGTCTCAATACCTTGGATTGCACTGGAACAT	2324
2107	Db	TCACTCTCTTCCAAAAGAGCCGCAATGAGTCTCAATACCTTGGATTGCACTGGAACAT	2166

QY 2325 GGAGACATCACTTTGAAGGATACAAATTTGTCCAAAGTCAGCGTTCGTGAGATCATTTCTG 2384  
DB |||||  
QY 2167 GGAGACATCACTTTGAAGGATACAAATTTGTCCAAAGTCAGCGTTCGTGAGATCATTTCTG 2226  
DB |||||  
QY 2385 ATGAACCTCAGAGCATGCTAAATTAATAAATCAAGCTATAATAACAGATGAAGAAACATGAC 2444  
DB |||||  
QY 2227 ATGAACCTCAGAGCATGCTAAATTAATAAATCAAGCTATAATAACAGATGAAGAAACATGAC 2286  
DB |||||  
QY 2445 AGTTTGGTGGCTCCACAGGAAACAGCTTCATATAAAGAGCATCTTGCCAAACAGCTTAGGA 2504  
DB |||||  
QY 2287 AGTTTGGTGGCTCCACAGGAAACAGCTTCATATAAAGAGCATCTTGCCAAACAGCTTAGGA 2346  
DB |||||  
QY 2505 GTGTCTGAAGATTGACAGAGTTGACCTTTCTCCAGCTGAGTGTAAGTGAATGCTCAT 2564  
DB |||||  
QY 2347 GTGTCTGAAGATTGACAGAGTTGACCTTTCTCCAGCTGAGTGTAAGTGAATGCTCAT 2406  
DB |||||  
QY 2565 GACAGGCTCAGATCCACCCCTGACCTTGAGACCCACAGCAGATTTAGAGTGGAACT 2624  
DB |||||  
QY 2407 GACAGGCTCAGATCCACCCCTGACCTTGAGACCCACAGCAGATTTAGAGTGGAACT 2466  
DB |||||  
QY 2625 CACACCCAAAACCATAGGCGGAAATGTGACAAAAGAAAGCCCCCATCTCTGATTGTT 2684  
DB |||||  
QY 2467 CACACCCAAAACCATAGGCGGAAATGTGACAAAAGAAAGCCCCCATCTCTGATTGTT 2526  
DB |||||  
QY 2685 CCCTGGAAGCCAGATGACAAAAGAAAGAAATCAACGGGAAAGAAAGAAAGAAAGCAGT 2744  
DB |||||  
QY 2527 CCCTGGAAGCCAGATGACAAAAGAAAGAAATCAACGGGAAAGAAAGAAAGAAAGCAGT 2586  
DB |||||  
QY 2745 AGAATGAGGAAATGCTGAAATACATAGCGCTTACTGAAGTGTACTTTGGAAGAAAG 2804  
DB |||||  
QY 2587 AGAATGAGGAAATGCTGAAATACATAGCGCTTACTGAAGTGTACTTTGGAAGAAAG 2646  
DB |||||  
QY 2805 CTGACGATTAACAGATAGTACTTGGGCTTTTGGCATTGGGAGAAAGAAAGATATTC 2864  
DB |||||  
QY 2647 CTGACGATTAACAGATAGTACTTGGGCTTTTGGCATTGGGAGAAAGAAAGATATTC 2706  
DB |||||  
QY 2865 CAGATCTTCGACGAGAGAGTCAATTGAAGACAAATTTGGCATACTTCACCTGATAGC 2924  
DB |||||  
QY 2707 CTAGATCTTCGACGAGAGAGTCAATTGAAGACAAATTTGGCATACTTCACCTGATAGC 2766  
DB |||||  
QY 2925 AAAAATCTGGGAGGCACTAAAAGATACATTTCCAGATATCCCTCAGATATGTAATAAA 2984  
DB |||||  
QY 2767 AAGAATCTGGGAGGCACTAAAAGATACATTTCCAGATATCCCTCAGATATGTAATAAA 2826  
DB |||||  
QY 2985 ATTCTAATAGCAAGTTTGGATTACATCGCGGAAAGTCCCTGCTACATGCTCCATG 3044  
DB |||||  
QY 2827 ATTCTAATAGCAAGTTTGGATTACATCGCGGAAAGTCCCTGCTACATGCTCCATG 2886  
DB |||||  
QY 3045 ATTGACCGATTGTTATCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGCAAGACG 3104  
DB |||||  
QY 2887 ATTGACCGATTGTTATCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGCAAGACG 2946  
DB |||||  
QY 3105 TCATTTCAAAAGTGGCCATTCTGAGGATATGAGTTTGGCTTCTCTTATTTTATAT 3164  
DB |||||  
QY 2947 TCATTTCAAAAGTGGCCATTCTGAGGATATGAGTTTGGCTTCTCTTATTTTATAT 3006  
DB |||||  
QY 3165 CTCATGAGTGCAGTGCAGCCACTGAATATCTCAAGTCTTTTGTATGAGTTGTATACAGAT 3224  
DB |||||  
QY 3007 CTCATGAGTGCAGTGCAGCCACTGAATATCTCAAGTCTTTTGTATGAGTTGTATACAGAT 3066  
DB |||||  
QY 3225 CAATCTGTCTGTCTGACAGAGAAATCCGAACACTGGCTACAGAAATTCAGAACTG 3284  
DB |||||  
QY 3067 CAATCTGTCTGTCTGACAGAGAAATCCGAACACTGGCTACAGAAATTCAGAACTG 3126  
DB |||||  
QY 3285 CCGTTAAGTTTGCAGGATTTGACAGGCTGGAACACATGCTTAATAATTCCTCAAAATG 3344  
DB |||||  
QY 3127 CCGTTAAGTTTGCAGGATTTGACAGGCTGGAACACATGCTTAATAATTCCTCAAAATG 3186  
DB |||||  
QY 3345 CTTCTGCTGATATCAGCAGCTAAATAATATTCACCAACTCAGGAATCCTACTATGAT 3404  
DB |||||  
QY 3187 CTTCTGCTGATATCAGCAGCTAAATAATATTCACCAACTCAGGAATCCTACTATGAT 3246  
DB |||||

QY 3405 CCCAACCTGCCACCGGTCACTAAAGTCTAGTAACAAACTGTAAACAGTAACCTGACAAA 3464  
DB |||||  
QY 3247 CCCAACCTGCCACCGGTCACTAAAGTCTAGTAACAAACTGTAAACAGTAACCTGACAAA 3306  
DB |||||  
QY 3465 ATCCCAAGCATATATAGGACAAACAAATATATAGTTTGAATCATCGGAGAGAGAA 3524  
DB |||||  
QY 3307 ATCCCAAGCATATATAGGACAAACAAATATATAGTTTGAATCATCGGAGAGAGAA 3366  
DB |||||  
QY 3525 ATCGCTTTTAAATGATTCTGACCAACGTTTCTCATGTGTGTGGCCAGTTGGATGACATA 3584  
DB |||||  
QY 3367 ATCGCTTTTAAATGATTCTGACCAACGTTTCTCATGTGTGTGGCCAGTTGGATGACATA 3426  
DB |||||  
QY 3585 AGAAAAACCTTAGAAGTTTGTTCCTGATGATGACCAATTCAGCAATCATANAAGT 3644  
DB |||||  
QY 3427 AGAAAAACCTTAGAAGTTTGTTCCTGATGATGACCAATTCAGCAATCATANAAGT 3486  
DB |||||  
QY 3645 GCTCAGACAGTGAAGCGCTTCTCAGGACATCTATGAATCCATGTTCCCATACCTTCC 3704  
DB |||||  
QY 3487 GCTCAGACAGTGAAGCGCTTCTCAGGACATCTATGAATCCATGTTCCCATACCTTCC 3546  
DB |||||  
QY 3705 CAATTTGAATCTGCCAAGAGATATCGAAACCGTTTCTTCTATATGATGAGCTGAGGAA 3764  
DB |||||  
QY 3547 CAATTTGAATCTGCCAAGAGATATCGAAACCGTTTCTTCTATATGATGAGCTGAGGAA 3606  
DB |||||  
QY 3765 TGGAGGCTTATCGAGACAAATTTGAAGTTTGGACCCATTTGTACTAGCAACATTTGATT 3824  
DB |||||  
QY 3607 TGGAGGCTTATCGAGACAAATTTGAAGTTTGGACCCATTTGTACTAGCAACATTTGATT 3666  
DB |||||  
QY 3825 ATGTTTACTATATCTCAATTTTGTCTGAGCAGTTTAAATGACCTTAAGCGGAATATTT 3884  
DB |||||  
QY 3667 ATGTTTACTATATCTCAATTTTGTCTGAGCAGTTTAAATGACCTTAAGCGGAATATTT 3726  
DB |||||  
QY 3885 CCAGAGGAGGATACACAAAGAGCTAGTCCCAATCGAATCAGATGATAGAGATC 3941  
DB |||||  
QY 3727 CCAGAGGAGGATACACAAAGAGCTAGTCCCAATCGAATCAGATGATAGAGATC 3783  
DB |||||

## RESULT 10

AR300472 3621 bp DNA linear PAT 12-JUN-2003  
LOCUS AR300472 Sequence 21 from patent US 6537785.

DEFINITION AR300472

ACCESSION AR300472

VERSION AR300472.1 GI:31687914

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE Unclassified.

AUTHORS 1 (bases 1 to 3621)

TITLE Canfield, W.M.

JOURNAL Methods of treating lysosomal storage diseases

Patent: US 6537785-A 21 25-MAR-2003;

FEATURES Location/Qualifiers

source 1..3621

/organism="unknown"

/mol\_type="genomic DNA"

## ORIGIN

Query Match 62.3%; Score 3485.4; DB 6; Length 3621;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 3492; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 297 AGCCGAGATCAATACCATGTTTGTGATTCCTATAGACAAATATTGCTGGAAGTCC 356

DB |||||

QY 115 AGCCGAGATCAATACCATGTTTGTGATTCCTATAGACAAATATTGCTGGAAGTCC 174

DB |||||

QY 357 TTTCAGAAATCGGCTTTGTCTGCCATGCGGATGACGTTGTTTACACCTGGGTGAATGGC 416

DB |||||

QY 175 TTTCAGAAATCGGCTTTGTCTGCCATGCGGATGACGTTGTTTACACCTGGGTGAATGGC 234

DB |||||

QY 417 ACAGATCTTTGAACACTACTGAAGAACTACAGCAGGTGAGAAACAGATGCGAGGAGCAG 476

DB |||||

QY 235 ACAGATCTTTGAACACTACTGAAGAACTACAGCAGGTGAGAAACAGATGCGAGGAGCAG 294

DB |||||

QY 477 AAAGCAATGAGAGAAATCCTTGGGAAAAACAACGGAACCTTACTAAGAAGAGTGAGAG 536  
DB 295 AAAGCAATGAGAGAAATCCTTGGGAAAAACAACGGAACCTTACTAAGAAGAGTGAGAG 354  
QY 537 CAGTTAGAGTGTTCCTAACACACTGCATTAAAGGTGCCAATGCTTGTACTGAGACCCAGCC 596  
DB 355 CAGTTAGAGTGTTCCTAACACACTGCATTAAAGGTGCCAATGCTTGTACTGAGACCCAGCC 414  
QY 597 CTGCCAGCAACATCACCTCTGAAGGAGCTGCCTCTTTATCTCTTTTCATTTCTGCTGCC 656  
DB 415 CTGCCAGCAACATCACCTCTGAAGGAGCTGCCTCTTTATCTCTTTTCATTTCTGCTGCC 474  
QY 657 AGTGACATTTTCAATGTTGCAAAACCAAAAAACCCCTTCTACCAATGTCTCAGTTGTTGTT 716  
DB 475 AGTGACATTTTCAATGTTGCAAAACCAAAAAACCCCTTCTACCAATGTCTCAGTTGTTGTT 534  
QY 717 TTTTGACAGTACTAAGAGTGTGAAGATGCCCACTCTGACCTGCTTTAAAGGAAATAGCAGA 776  
DB 535 TTTTGACAGTACTAAGAGTGTGAAGATGCCCACTCTGACCTGCTTTAAAGGAAATAGCAGA 594  
QY 777 CAGACAGTATGAGGGGGTACTTTGACACAGATAAAGTCCCTGGATTAGTGCTAATG 836  
DB 595 CAGACAGTATGAGGGGGTACTTTGACACAGATAAAGTCCCTGGATTAGTGCTAATG 654  
QY 837 CAGATTTTGGCTTTCTGAGTGGATTTCCACCAACATTCAGGAAACAAATCAACTAAAA 896  
DB 655 CAGATTTTGGCTTTCTGAGTGGATTTCCACCAACATTCAGGAAACAAATCAACTAAAA 714  
QY 897 ACATAATGCCAGAAATCTTTCTCTAAAGTCAAACTGTTGCAAGTGTCTGAGGCCC 956  
DB 715 ACATAATGCCAGAAATCTTTCTCTAAAGTCAAACTGTTGCAAGTGTCTGAGGCCC 774  
QY 957 AGTGTAGGCTTTCTAAACTGTAATAACCCCAAGGATTTTCAAGAAATGAATAAGCAAACT 1016  
DB 775 AGTGTAGGCTTTCTAAACTGTAATAACCCCAAGGATTTTCAAGAAATGAATAAGCAAACT 834  
QY 1017 AAGAGACATGACCAATGATGGAAGAAACATGACCAATAGTCCCTGCAATTTATTATGG 1076  
DB 835 AAGAGACATGACCAATGATGGAAGAAACATGACCAATAGTCCCTGCAATTTATTATGG 894  
QY 1077 GATCTGAGGCGCCATCAGCCAGTCTTAAGCAGGATGAAGACATCTCTGCCAGTCTGTTGAA 1136  
DB 895 GATCTGAGGCGCCATCAGCCAGTCTTAAGCAGGATGAAGACATCTCTGCCAGTCTGTTGAA 954  
QY 1137 GATAACGAAGAACTGAGGTACTCATTTGGGATCTATCGAGAGGATGCAACATGCTGCG 1196  
DB 955 GATAACGAAGAACTGAGGTACTCATTTGGGATCTATCGAGAGGATGCAACATGCTGCG 1014  
QY 1197 AATATTTTCTGTCACCAAGGCGAGATTCCTGCTGCAACCTTTGACAACTCTCGA 1256  
DB 1015 AATATTTTCTGTCACCAAGGCGAGATTCCTGCTGCAACCTTTGACAACTCTCGA 1074  
QY 1257 GTGACAAATAGTAACACACAGGATGTTTTCGAAAATTTGAGCCACTTCCCTACCTTTAGT 1316  
DB 1075 GTGACAAATAGTAACACACAGGATGTTTTCGAAAATTTGAGCCACTTCCCTACCTTTAGT 1134  
QY 1317 TCACCTGCTATTGAAAGTACATTCATCGCATCGAAGGCTGTCCAGAAATTTATTATAC 1376  
DB 1135 TCACCTGCTATTGAAAGTACATTCATCGCATCGAAGGCTGTCCAGAAATTTATTATAC 1194  
QY 1377 CTAATATGATGATGCTCATGTTTGGGAAAGATGCTGCGCCAGATGATTTTACAGTCACTCC 1436  
DB 1195 CTAATATGATGATGCTCATGTTTGGGAAAGATGCTGCGCCAGATGATTTTACAGTCACTCC 1254  
QY 1437 AAAGCCGAAGAGTTTATTTGACATGGCTGTGCAAACTGTGCGGAGGCTGCGCCAGGT 1496  
DB 1255 AAAGCCGAAGAGTTTATTTGACATGGCTGTGCAAACTGTGCGGAGGCTGCGCCAGGT 1314  
QY 1497 TCCTGATTTAAGGATGGCTATTGTGACAGGCTTCTTAATAATTACGCTGCGATTGGGAT 1556  
DB 1315 TCCTGATTTAAGGATGGCTATTGTGACAGGCTTCTTAATAATTACGCTGCGATTGGGAT 1374  
QY 1557 GGTGGGGATGCTCTGGAACACAGTGGAGGAGTGCCTATATTGACAGGAGTGGAGGTACT 1616

DB 1375 GGTGGGGATGCTCTGGAACACAGTGGAGGAGTGCCTATATTGACAGGAGTGGAGGTACT 1434  
QY 1617 GGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTCTTAC 1676  
DB 1435 GGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTCTTAC 1494  
QY 1677 TGTAAATCAGAGATGTGCGAATTCCTGGCTCGCTGATAAGTTCTGTGACCAAGCATGCAAT 1736  
DB 1495 TGTAAATCAGAGATGTGCGAATTCCTGGCTCGCTGATAAGTTCTGTGACCAAGCATGCAAT 1554  
QY 1737 GTCTGTGCTGTGGGTTTGATGTGCGAGTGTGGGCAAGATCAATTTTCATGAATTTGTAT 1796  
DB 1555 GTCTGTGCTGTGGGTTTGATGTGCGAGTGTGGGCAAGATCAATTTTCATGAATTTGTAT 1614  
QY 1797 AAGTGTATCTCTCCCAACACAGACTCACTATATTATTCCTCAAAAGGTGAATGCCCTGCC 1856  
DB 1615 AAGTGTATCTCTCCCAACACAGACTCACTATATTATTCCTCAAAAGGTGAATGCCCTGCC 1674  
QY 1857 TATTTTCAGCTTTTCAGAAAGTAGCAAAAGAGGAGTGAAGGTCCTATAGTACAAATCCA 1916  
DB 1675 TATTTTCAGCTTTTCAGAAAGTAGCAAAAGAGGAGTTGAAGTGCCTATAGTACAAATCCA 1734  
QY 1917 ATAATTCGACATGCTTCTATTTGCCAACAGTGGAAACCAATCCACCTCATATATGACAGT 1976  
DB 1735 ATAATTCGACATGCTTCTATTTGCCAACAGTGGAAACCAATCCACCTCATATATGACAGT 1794  
QY 1977 GGAATGAATGCCACCAATACATTTTAACTCACAAGTTTCAAAATACAAACGATGAAG 2036  
DB 1795 GGAATGAATGCCACCAATACATTTTAACTCACAAGTTTCAAAATACAAACGATGAAG 1854  
QY 2037 TCCAAATGCGAGATAACAGTGGAGTGGACACAAGGAGGAGCCAAACCTGAATTTCTACG 2096  
DB 1855 TCCAAATGCGAGATAACAGTGGAGTGGACACAAGGAGGAGCCAAACCTGAATTTCTACG 1914  
QY 2097 GCCCAGAGGAGTTTACGAAATTTAGTTAGTCCCAATACACTTCTCCAGAGGCGGAAATC 2156  
DB 1915 GCCCAGAGGAGTTTACGAAATTTAGTTAGTCCCAATACACTTCTCCAGAGGCGGAAATC 1974  
QY 2157 CTTTTCAGAGATATTCCTCAAAAGAAAAACGCTTCCCGAAGTTTAAAGAGACATGATTTAAC 2216  
DB 1975 CTTTTCAGAGATATTCCTCAAAAGAAAAACGCTTCCCGAAGTTTAAAGAGACATGATTTAAC 2034  
QY 2217 TCACAAAGGAGAGCCACAGGAGGAGGTAATAATTCCTCCCTGGTAAATATTTTCACTCTTCCA 2276  
DB 2035 TCACAAAGGAGAGCCACAGGAGGAGGTAATAATTCCTCCCTGGTAAATATTTTCACTCTTCCA 2094  
QY 2277 AAAGAGCCCCAGTTGAGTCTCAATACCTTGGATTTCGAACTGGAAACATGGAGACATCACT 2336  
DB 2095 AAAGAGCCCCAGTTGAGTCTCAATACCTTGGATTTCGAACTGGAAACATGGAGACATCACT 2154  
QY 2337 TTGAAAGGATACAAATTTGTCCAAAGTCAAGCTTGTCTGAGATCAATTTCTGATGAACCTCACAG 2396  
DB 2155 TTGAAAGGATACAAATTTGTCCAAAGTCAAGCTTGTCTGAGATCAATTTCTGATGAACCTCACAG 2214  
QY 2397 CATGCTAAATAAATAAATAAAGCTATAATAACAGATGAAACCAATGACAGTTTGTGGCT 2456  
DB 2215 CATGCTAAATAAATAAATAAAGCTATAATAACAGATGAAACCAATGACAGTTTGTGGCT 2374  
QY 2457 CCACAGAAAAACAGGTTTCATAAAAGCATCTTCCCAACAGCTTAGGAGTGTCTGAAAGA 2516  
DB 2275 CCACAGAAAAACAGGTTTCATAAAAGCATCTTCCCAACAGCTTAGGAGTGTCTGAAAGA 2334  
QY 2517 TTGACAGGTTGACTTTTCTGCAAGTGTGTAAGTGAATGATGATGATGATGATGATGATGATGAT 2576  
DB 2335 TTGACAGGTTGACTTTTCTGCAAGTGTGTAAGTGAATGATGATGATGATGATGATGATGATGAT 2394  
QY 2577 AATCCACCCCTGGAGTTGGAGACCAAGCAAGATTTTAGAGTGGAAACTCACACCCAAAAA 2636  
DB 2395 AATCCACCCCTGGAGTTGGAGACCAAGCAAGATTTTAGAGTGGAAACTCACACCCAAAAA 2454  
QY 2637 ACCATAGGCGGAATGTGACAAAGAAAGCCCCCATCTCTGATTTTCCATCTGGAAGC 2696

2455	ACCATAGCGCGAATGTGACAAAAGAAAAGCCCCCATCTCTGTGATTGTTCACTGGAAAGC	2514
2497	CAGATGACAAAAGAAAAAATAATCACAGGGAAGAAAAGAGAACAGTAGAATGGAGGAA	2756
2515	CAGATGACAAAAGAAAAAATAATCACAGGGAAGAAAAGAGAACAGTAGAATGGAGGAA	2574
2757	AATCCTGAAATACATAGGCGTTACTCGAAGTGTACTTGGAGAAAGCTGCGACATTAC	2816
2575	AATCCTGAAATACATAGGCGTTACTCGAAGTGTACTTGGAGAAAGCTGCGACATTAC	2634
2817	ACAGATAGTTACTTTGGGCTTTTGGCCATGGGAGAAAAAAGTATTTTCCAAAGATCTTCTC	2876
2635	ACAGATAGTTACTTTGGGCTTTTGGCCATGGGAGAAAAAAGTATTTTCTTAGATCTTCTC	2694
2877	GACGAAGAAGAGTCATTGAAGACACAATTTGGCATACTTCTAGTAGACAAAATACTGGG	2936
2695	GACGAAGAAGAGTCATTGAAGACACAATTTGGCATACTTCTAGTAGACAAAATACTGGG	2754
2937	AGGCAACTAAAGATACATTTGCGAGTTCCCTCAGATATGTAAATAAAATTTCTAAATAGC	2996
2755	AGGCAACTAAAGATACATTTGCGAGTTCCCTCAGATATGTAAATAAAATTTCTAAATAGC	2814
2997	AAGTTTGGATTACATCGCGAAAAGTCCCTGCTCACATGCCTCACATGATGACCGGATT	3056
2815	AAGTTTGGATTACATCGCGAAAAGTCCCTGCTCACATGCCTCACATGATGACCGGATT	2874
3057	GTTATGCAAGAACTGCAAGATATGTTCCCTGGAAGAAATTTGACAAGACGTCATTTACAAA	3116
2875	GTTATGCAAGAACTGCAAGATATGTTCCCTGGAAGAAATTTGACAAGACGTCATTTACAAA	2934
3117	GTGCGCCATTCGAGGATATGCAAGTTGCGCTTCCTTATTTTTATTATCTCATGAGTGCA	3176
2935	GTGCGCCATTCGAGGATATGCAAGTTGCGCTTCCTTATTTTTATTATCTCATGAGTGCA	2994
3177	GTGCGCCACCTGAATATATCTCAAGTCCTTTGATGAAAGTTGATACAGATCAATCTGCTGC	3236
2995	GTGCGCCACCTGAATATATCTCAAGTCCTTTGATGAAAGTTGATACAGATCAATCTGCTGC	3054
3237	TTGTCGTACACAGAAAATCCGAACACTGGCTACCAAGAAATTCACGAACCTGCCGTTAAGTTG	3296
3055	TTGTCGTACACAGAAAATCCGAACACTGGCTACCAAGAAATTCACGAACCTGCCGTTAAGTTG	3114
3297	CAGGATTGACAGGCTCTGGAACACATGCTAATAAATTTGCTCAAAAATGCTTCTGCTGAT	3356
3115	CAGGATTGACAGGCTCTGGAACACATGCTAATAAATTTGCTCAAAAATGCTTCTGCTGAT	3174
3357	ATCAGCGAGCTAAATATATTTCCACCAACTCAGGAATCCTACTATGATCCCAACCTGCCA	3416
3175	ATCAGCGAGCTAAATATATTTCCACCAACTCAGGAATCCTACTATGATCCCAACCTGCCA	3234
3417	CCGGTCACTAAAGTCTAGTAAACAAACCTGTAAACCGAGTAACTGCAAAAAATCCACAAGCA	3476
3235	CCGGTCACTAAAGTCTAGTAAACAAACCTGTAAACCGAGTAACTGCAAAAAATCCACAAGCA	3294
3477	TATAAGGACAAAAACAATATAGTTTGAATATCATGGGAGAGAGAAATCGCTTTTAAA	3536
3295	TATAAGGACAAAAACAATATAGTTTGAATATCATGGGAGAGAGAAATCGCTTTTAAA	3354
3537	ATGATTGCTACCAACGTTTCTCATGTGGTTGGCCAGTTGGATGACATAAGAAAAAACCCCT	3596
3355	ATGATTGCTACCAACGTTTCTCATGTGGTTGGCCAGTTGGATGACATAAGAAAAAACCCCT	3414
3597	AGGAAGTTTGTTCGCTGAATGACAAATGACCACAATCATAAAGATGCTTCAGACAGTG	3656
3415	AGGAAGTTTGTTCGCTGAATGACAAATGACCACAATCATAAAGATGCTTCAGACAGTG	3474
3657	AAGGCTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCACTACCTTCCCAATTTGAACTG	3716
3475	AAGGCTGTTCTCAGGAGCTTCTATGAATCCATGTTTCCCACTACCTTCCCAATTTGAACTG	3534
3717	CCAAAGAGAGTATCGAAAACCGTTTCTCCTTCATATGATGAGCTGCGAGGAATGGAGGCTTAT	3776
3535	CCAAAGAGAGTATCGAAAACCGTTTCTCCTTCATATGATGAGCTGCGAGGAATGGAGGCTTAT	3594

Oy	3777	CGAGCAAAATTGAAGTTTTCGAC	3799
Db	3595	CGAGCAAAATTGAAGTAGTAGTC	3617

  

RESULT	11		
LOCUS	AR428749	3621 bp	DNA
DEFINITION	Sequence 21 from patent US 6642038.		
ACCESSION	AR428749		
VERSION	AR428749.1	GI:40188483	
KEYWORDS	.		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 3621)		
TITLE	Canfield,W.M.		
JOURNAL	GlcNac phosphotransferase of the lysosomal targeting pathway		
FEATURES	Patent: US 6642038-A 21-04-NOV-2003;		
	Location/Qualifiers		
	1..3621		
	/organism="unknown"		
	/mol type="genomic DNA"		
	source		

Query Match	62.3%;	Score 3485.4;	DB 6;	Length 3621;
Best Local Similarity	99.7%;	Pred. No. 0;		
Matches 3492;	Conservative 0;	Mismatches 11;	Indels 0;	Gaps 0;
QY	297	AGCGAGATCAATACCACTGTTTGTGTTGATTCCTATAGACACAATATGCTGGAAAGTCC	356	
Db	115	AGCGAGATCAATACCACTGTTGTGTTGATTCCTATAGACACAATATGCTGGAAAGTCC	174	
QY	357	TTTCAGAAATCGGCTTTGTCGCCCACGCCGATTGACCGTTGTTTACACTGGGTGAATGGC	416	
Db	175	TTTCAGAAATCGGCTTTGTCGCCCACGCCGATTGACCGTTGTTTACACTGGGTGAATGGC	234	
QY	417	ACAGATCTCTCAACTACTGTAAGGAACCTACACAGGTGAGAGAACAGATGGAAGGAGGACAG	476	
Db	235	ACAGATCTCTGAACTACTGTAAGGAACCTACACAGGTGAGAGAACAGATGGAAGGAGGACAG	294	
QY	477	AAAGCAATGAGAGAAATCTTTGGGAAAAACACAAACCGAACCTTACTAAGAAAGATGAGAAG	536	
Db	295	AAAGCAATGAGAGAAATCTTTGGGAAAAACACAAACCGAACCTTACTAAGAAAGATGAGAAG	354	
QY	537	CAGTTAGAGTGTGTTGCTAAACACACTGCAATTAAGGTGCCAATGCTTGTTACTGGACCCAGCC	596	
Db	355	CAGTTAGAGTGTGTTGCTAAACACACTGCAATTAAGGTGCCAATGCTTGTTACTGGACCCAGCC	414	
QY	597	CTGCCAGCCAAACATCACCCCTGAAGGAGCGTGCCATCTCTTTATCCTCTCTTTTCATTTCTGCC	656	
Db	415	CTGCCAGCCAAACATCACCCCTGAAGGAGCGTGCCATCTCTTTATCCTCTCTTTTCATTTCTGCC	474	
QY	657	AGTGACATTTTCAATGTTTGCAAAACCCAAACAAACCCCTCTCAACCAATGCTCAGTGTGTTGTT	716	
Db	475	AGTGACATTTTCAATGTTTGCAAAACCCAAACAAACCCCTCTCAACCAATGCTCAGTGTGTTGTT	534	
QY	717	TTTGACAGTACTAAGGATGTTGAAGATGCCCACCTCTCGACTGCTTTAAAGGAAATATGACAGA	776	
Db	535	TTTGACAGTACTAAGGATGTTGAAGATGCCCACCTCTGCACTGCTTTAAGGAAATATGACAGA	594	
QY	777	CAGACAGTATGGAGGGGTACTTTGACAACACAGATAAAGAAATGCCCTCGAATTAGTGTCTAATG	836	
Db	595	CAGACAGTATGGAGGGGCTACTTTGACAACACAGATAAAGAAAGTCCCTCGAATTAGTGTCTAATG	654	
QY	837	CAAGATTTGGCTTTCCTGAGTGGATTTCCACCACCACTTCAAGGAAAAACAAATCAACTAAAA	896	
Db	655	CAAGATTTGGCTTTCCTGAGTGGATTTCCACCACCACTTCAAGGAAAAACAAATCAACTAAAA	714	
QY	897	ACAAAATTGCCAGAAAAATCTTTCTCTTAAAGTCAAACTGTTGAGTGTGTTATTCAGAGGCC	956	
Db	715	ACAAAATTGCCAGAAAAATCTTTCTCTTAAAGTCAAACTGTTGAGTGTGTTATTCAGAGGCC	774	

QY 957 AGTGTAGCGCTTCTAAACTGATATAACCCCAAGGATTTTCAAGATTAAGCAAACT 1016  
DB 775 AGTGTAGCGCTTCTAAACTGATATAACCCCAAGGATTTTCAAGATTAAGCAAACT 834  
QY 1017 AAGAAAGAACATGACCAATGATGAAAGAACGACATAAGCTCCGTCATATTATATGG 1076  
DB 835 AAGAAAGAACATGACCAATGATGAAAGAACGACATAAGCTCCGTCATATTATATGG 894  
QY 1077 GATCTGAGCGGCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCGCAGTCGTTTGAA 1136  
DB 895 GATCTGAGCGGCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCGCAGTCGTTTGAA 954  
QY 1137 GATAACGAAGAACTCAGGTACTCATTTGGGATCTATCGAGAGCATGCACCATGGGTCGG 1196  
DB 955 GATAACGAAGAACTCAGGTACTCATTTGGGATCTATCGAGAGCATGCACCATGGGTCGG 1014  
QY 1197 AATATTTTCAATGTCACCAACGGGAGATTCATCTCGGTGAACTTGAACATCTCTCGA 1256  
DB 1015 AATATTTTCAATGTCACCAACGGGAGATTCATCTCGGTGAACTTGAACATCTCTCGA 1074  
QY 1257 GTGACAAATAGTAACACACAGGATGTTTTGGAATTTGAGCAATTTGAGCTTACCTTTAGT 1316  
DB 1075 GTGACAAATAGTAACACACAGGATGTTTTGGAATTTGAGCAATTTGAGCTTACCTTTAGT 1134  
QY 1317 TCACCTGCTATTTGAAAGTCACATTCATCGCATCGAAGGCTGTCCGAGAGTTTATTTAC 1376  
DB 1135 TCACCTGCTATTTGAAAGTCACGTTTCACTCGCATCGAAGGCTGTCCGAGAGTTTATTTAC 1194  
QY 1377 CTAATATGATGTCATGTTGGGAAGATGTCGCGCAGATGATTTTTACAGTCACTCC 1436  
DB 1195 CTAATATGATGTCATGTTGGGAAGATGTCGCGCAGATGATTTTTACAGTCACTCC 1254  
QY 1437 AAGGCCAGAAAGTTTATTTGACATGGCTGTGCGCAAACTGTGCGAGGGTGCACAGT 1496  
DB 1255 AAGGCCAGAAAGTTTATTTGACATGGCTGTGCGCAAACTGTGCGAGGGTGCACAGT 1314  
QY 1497 TCCTGGATTAAGGATGGCTATTTGACAAAGCTTCTAATAATTCAGCTCGCATTTGGAT 1556  
DB 1315 TCCTGGATTAAGGATGGCTATTTGACAAAGCTTCTAATAATTCAGCTCGCATTTGGAT 1374  
QY 1557 GGTGGGATGCTCTGGAACAGTGGAGGATGCTCTATATTCGAGGAGTGGAGTACT 1616  
DB 1375 GGTGGGATGCTCTGGAACAGTGGAGGATGCTCTATATTCGAGGAGTGGAGTACT 1434  
QY 1617 GGGATATTTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTCTTAC 1676  
DB 1435 GGGATATTTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTCTTAC 1494  
QY 1677 TGTAATCAGGATGTCGAAATTCCTGGCTCGCTGATTAAGTCTGTGACCAAGCATGCAT 1736  
DB 1495 TGTAATCAGGATGTCGAAATTCCTGGCTCGCTGATTAAGTCTGTGACCAAGCATGCAT 1554  
QY 1737 GTCTTGTCTGTGGGTTTGTGCTGGGACCTGTGGGCAAGATCATTTTCATGAATTTGTAT 1796  
DB 1555 GTCTTGTCTGTGGGTTTGTGCTGGGACCTGTGGGCAAGATCATTTTCATGAATTTGTAT 1614  
QY 1797 AAAGTGATCTTCTCCCAACACAGATCTCATATATTTATTCGAAAGGTTGCTGCTGCT 1856  
DB 1615 AAAGTGATCTTCTCCCAACACAGATCTCATATATTTATTCGAAAGGTTGCTGCTGCT 1674  
QY 1857 TATTTTCAGCTTTGCAAGATGACCAAAAGAGAGTTGAAGTGGCTATAGTGACAAATCCA 1916  
DB 1675 TATTTTCAGCTTTGCAAGATGACCAAAAGAGAGTTGAAGTGGCTATAGTGACAAATCCA 1734  
QY 1917 ATAAATTCAGATGCTTCTATTTGCCAACAGTGGAAACCATCCACCTCATATGACAGT 1976  
DB 1735 ATAAATTCAGATGCTTCTATTTGCCAACAGTGGAAACCATCCACCTCATATGACAGT 1794  
QY 1977 GGAATGAATGCCACCAATACATTTTAACTCTCAGCTTTCAAATAACAAAGATGAAG 2036  
DB 1795 GGAATGAATGCCACCAATACATTTTAACTCTCAGCTTTCAAATAACAAAGATGAAG 1854

QY 2037 TTTCAAAATGCAATACAGTGGAGTGGACACAAAGGGAGGACAAAACCTGAATTTCTACG 2096  
DB 1855 TTTCAAAATGCAATACAGTGGAGTGGACACAAAGGGAGGACAAAACCTGAATTTCTACG 1914  
QY 2097 GCCCAGAGGGTTACGAAAATTTAGTTAGTCCCATAACTTCTTCCAGAGGCGGAAATC 2156  
DB 1915 GCCCAGAGGGTTACGAAAATTTAGTTAGTCCCATAACTTCTTCCAGAGGCGGAAATC 1974  
QY 2157 CTTTTCAGAGATATTCCTCAAGAAAACCGTTCCTCCGAAAGTTTAAGAGACATGATGTAAC 2216  
DB 1975 CTTTTCAGAGATATTCCTCAAGAAAACCGTTCCTCCGAAAGTTTAAGAGACATGATGTAAC 2034  
QY 2217 TCAACAGGAGAGCCACAGGAAGAGGTGAAAATTTCCCTCGTAAATATTTTCACTCTTCCA 2276  
DB 2035 TCAACAGGAGAGCCACAGGAAGAGGTGAAAATTTCCCTCGTAAATATTTTCACTCTTCCA 2094  
QY 2277 AAGAGCCCCAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAACATGGAGACATCACT 2336  
DB 2095 AAGAGCCCCAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAACATGGAGACATCACT 2154  
QY 2337 TTGAAGGATACAAATTTGTCCTCAAGTCCAGCTTCTGAGATCAATTTCTGATGAACCTCACAG 2396  
DB 2155 TTGAAGGATACAAATTTGTCCTCAAGTCCAGCTTCTGAGATCAATTTCTGATGAACCTCACAG 2214  
QY 2397 CATGCTAAAAATAAAAAATCAAGCTATAATAACAGATGAAAACAAATGACAGTTTGGTGGCT 2456  
DB 2215 CATGCTAAAAATAAAAAATCAAGCTATAATAACAGATGAAAACAAATGACAGTTTGGTGGCT 2274  
QY 2457 CCACAGGAAAAACAGGTTTCAATAAAGCATCTTGCCAAAACAGCTTAGAGTGTCTGAAGA 2516  
DB 2275 CCACAGGAAAAACAGGTTTCAATAAAGCATCTTGCCAAAACAGCTTAGAGTGTCTGAAGA 2334  
QY 2517 TTGCAGAGGTTGATTTTCTGTCAGTGTGTAATAAGTGAATGCTCATGACCAAGGCTCAG 2576  
DB 2335 TTGCAGAGGTTGATTTTCTGTCAGTGTGTAATAAGTGAATGCTCATGACCAAGGCTCAG 2394  
QY 2577 AATCCACCCCTGGACCTTGAGACCAACAGAAAGTTTAGAGTGGAAACTCACACCCAAAAA 2636  
DB 2395 AATCCACCCCTGGACCTTGAGACCAACAGAAAGTTTAGAGTGGAAACTCACACCCAAAAA 2454  
QY 2637 ACCATAGGCGGAAATGTGACAAAGAAAAGCCCATCTCTGATGTTCCACCTGGAAAGC 2696  
DB 2455 ACCATAGGCGGAAATGTGACAAAGAAAAGCCCATCTCTGATGTTCCACCTGGAAAGC 2514  
QY 2697 CAGATGACAAAAGAAAAGAAATCACAGGGAAGAAAAGAGAACAGTAGTAGAGGAA 2756  
DB 2515 CAGATGACAAAAGAAAAGAAATCACAGGGAAGAAAAGAGAACAGTAGTAGAGGAA 2574  
QY 2757 AATGCTGAATACATAGGCTTACTGAAGTGTACTTGGAGAGAGCTGACAGCATAC 2816  
DB 2575 AATGCTGAATACATAGGCTTACTGAAGTGTACTTGGAGAGAGCTGACAGCATAC 2634  
QY 2817 ACAGATAGTTTACTTGGGCTTTTGGCCATGGGAAAAAAAAGTATTTTCCAAAGATCTTCTC 2876  
DB 2635 ACAGATAGTTTACTTGGGCTTTTGGCCATGGGAAAAAAAAGTATTTTCTAGATCTTCTC 2694  
QY 2877 GAGAGAGAGATCATTTGAAGACACAAATTTGGCCTACTTCTGATAGAGAAAATCTGGG 2936  
DB 2695 GAGCAGAGAGATCATTTGAAGACACAAATTTGGCCTACTTCTGATAGAGAAAATCTGGG 2754  
QY 2937 AGCACAATAAGATACATTTGAGATTCCTCAGATATGTAATAAATAATTTCTAAATAGC 2996  
DB 2755 AGCACAATAAGATACATTTGAGATTCCTCAGATATGTAATAAATAATTTCTAAATAGC 2814  
QY 2997 AAGTTTGGATTCATCGGGGAAAGTCCCTGCTCAGATGATGATGATGATGATGATGATGAT 3056  
DB 2815 AAGTTTGGATTCATCGGGGAAAGTCCCTGCTCAGATGATGATGATGATGATGATGATGAT 2874  
QY 3057 GTTATCCAGAGCTCAAGATATGTTCCCTGAGAAATTTGACAAAGAGCTCATTTTCACAAA 3116  
DB 2875 GTTATCCAGAGCTCAAGATATGTTCCCTGAGAAATTTGACAAAGAGCTCATTTTCACAAA 2934  
QY 3117 GTGGCGCAATCTGAGGATATGCAAGTTTGGCTTCTCTTATTTTATTTATCTCATGATGCA 3176



Db	2935	GTGGCCCATCTGAGGATATGCACTTGGCTTCTCTATTTATATCATGATGCA	2994	Db	115	AGCCGAGATCAATACCATGTTTGTTCATTCCTATAGACAATATGCTGGAAGTCC	174
Qy	3177	GTGACGACCTGAATATATCTCAAGTCTTTGATGAAGTGTATACAGATCAATCTGSGTGC	3236	Qy	357	TTTCAGAAATCGGCTTGTCTGCCATGCGATGAGCTTGTATACCTGGGTGAATGGC	416
Db	2995	GTGACGACCTGAATATATCTCAAGTCTTTGATGAAGTGTATACAGATCAATCTGSGTGC	3054	Db	175	TTTCAGAAATCGGCTTGTCTGCCATGCGATGAGCTTGTATACCTGGGTGAATGGC	234
Qy	3237	TTGTCTGACAGAGAAATCCGAACTGCTGCTACAGAAATTCACGAACTGCGCTTAAGTTG	3296	Qy	417	ACAGATCTTGACTACTCAAGGAACTACAGAGGTCAGAGAAAGATGAGAGAGAGCAG	476
Db	3055	TTGTCTGACAGAGAAATCCGAACTGCTGCTACAGAAATTCACGAACTGCGCTTAAGTTG	3114	Db	235	ACAGATCTTGACTACTCAAGGAACTACAGAGGTCAGAGAAAGATGAGAGAGAGCAG	294
Qy	3297	CAGGATTTGACAGGCTGGAACACATGCTAATAAATTCCTCAAAATTCCTCTGCTGAT	3356	Qy	477	AAAGCAATGAGAGAAATCTTTGGGAAACACAAAGGAACTTACTAAGAGAGTGAAG	536
Db	3115	CAGGATTTGACAGGCTGGAACACATGCTAATAAATTCCTCAAAATTCCTCTGCTGAT	3174	Db	295	AAAGCAATGAGAGAAATCTTTGGGAAACACAAAGGAACTTACTAAGAGAGTGAAG	354
Qy	3357	ATCAGCGAGCTAATAATATTCACCAACTCAGGAATCTTACTATGATCCCAACCTGCCA	3416	Qy	537	CAGTTAGAGTGTTCCTAACACACTGCAATTAAGGTGCCAATGCTTGTCTGACCCAGCC	596
Db	3175	ATCAGCGAGCTAATAATATTCACCAACTCAGGAATCTTACTATGATCCCAACCTGCCA	3234	Db	355	CAGTTAGAGTGTTCCTAACACACTGCAATTAAGGTGCCAATGCTTGTCTGACCCAGCC	414
Qy	3417	CGGTCACTAAAGTCTAGTAACAACTGTAAACAGTAAGTGAACAAATCCCAAGCA	3476	Qy	597	CTGCCAGCAACATCACTCCCTGAAGGACGTGCCATCTCTTTATCTCTTTTCACTTCTGCC	656
Db	3235	CGGTCACTAAAGTCTAGTAACAACTGTAAACAGTAAGTGAACAAATCCCAAGCA	3294	Db	415	CTGCCAGCAACATCACTCCCTGAAGGACGTGCCATCTCTTTATCTCTTTTCACTTCTGCC	474
Qy	3477	TATAAGGACAAACAAATATAGTTTGAATCATGGAAGAAAGAAATCGCTTTTAA	3536	Qy	657	AGTGACATTTTCAATGTTGCAAAACCAAAACCTTCTTCAATGCTCTCAGTTGTTGT	716
Db	3295	TATAAGGACAAACAAATATAGTTTGAATCATGGAAGAAAGAAATCGCTTTTAA	3354	Db	475	AGTGACATTTTCAATGTTGCAAAACCAAAACCTTCTTCAATGCTCTCAGTTGTTGT	534
Qy	3537	ATGATTCGACCAAGTTTCTCATGTGTTGGCCAGTTGGATGACATAGAAAGAAACCT	3596	Qy	717	TTTGCAGTACTTAAGGATGTTGAAGATGCCACCTCTGGACTGCTTAAAGGAAATAGCAGA	776
Db	3355	ATGATTCGACCAAGTTTCTCATGTGTTGGCCAGTTGGATGACATAGAAAGAAACCT	3414	Db	535	TTTGCAGTACTTAAGGATGTTGAAGATGCCACCTCTGGACTGCTTAAAGGAAATAGCAGA	594
Qy	3597	AGGAAGTTTGTTCCTGCTGAATGACAACTTACCAACATCATTAAGATGCTCAGACAGTG	3656	Qy	777	CAGACAGTATGAGGGGTACTTGACAAAGATTAAGAAAGTCCCTGAGTATAGTCTAATG	836
Db	3415	AGGAAGTTTGTTCCTGCTGAATGACAACTTACCAACATCATTAAGATGCTCAGACAGTG	3474	Db	595	CAGACAGTATGAGGGGTACTTGACAAAGATTAAGAAAGTCCCTGAGTATAGTCTAATG	654
Qy	3657	AAGGCTGTTCTCAGGACTTCTATGAATCCATGTTCCCATACCTTCCCATTTGACTG	3716	Qy	837	CAAGATTTTGGCTTCTCAGTGGATTTCCACCAACTTCAAGGAAACAAATCAACTAAA	896
Db	3475	AAGGCTGTTCTCAGGACTTCTATGAATCCATGTTCCCATACCTTCCCATTTGACTG	3534	Db	655	CAAGATTTTGGCTTCTCAGTGGATTTCCACCAACTTCAAGGAAACAAATCAACTAAA	714
Qy	3717	CCAGAGAGTATCGAAACCGTTTCTTCATATGATGATGCTGAGGAAATGAGGGCTTAT	3776	Qy	897	ACAAATTTGCCAGAAAAATCTTTCTCTAAAGTCAAACTGTTGCAAGTGTATTCAGAGGCC	956
Db	3535	CCAGAGAGTATCGAAACCGTTTCTTCATATGATGATGCTGAGGAAATGAGGGCTTAT	3594	Db	715	ACAAATTTGCCAGAAAAATCTTTCTCTAAAGTCAAACTGTTGCAAGTGTATTCAGAGGCC	774
Qy	3777	CGAGACAAATGAAAGTTTGGAC	3799	Qy	957	AGTGTAGGCTTCTAAAATGNAATACCCAGGATTTTCAAGATTTGAATTAAGCAAACT	1016
Db	3595	CGAGACAAATGAAAGTTTGGAC	3617	Db	775	AGTGTAGGCTTCTAAAATGNAATACCCAGGATTTTCAAGATTTGAATTAAGCAAACT	834
RESULT 12				Qy	1017	AAGAAGAACATGACCATTTGATGAAAAAGAACTGACCAATGCTCTGCATATTTATTTGG	1076
AR442838				Db	835	AAGAAGAACATGACCATTTGATGAAAAAGAACTGACCAATGCTCTGCATATTTATTTGG	894
LOCUS	AR442838	3621 bp	DNA	linear	PAT	20-FEB-2004	
DEFINITION	Sequence 21 from patent US 6670165.						
ACCESSION	AR442838						
VERSION	AR442838.1	GI:42670314					
KEYWORDS	Unknown.						
SOURCE	Unknown.						
ORGANISM	Unclassified.						
REFERENCE	1 (bases 1 to 3621)						
AUTHORS	Canfield, W.M.						
TITLE	Methods for producing highly phosphorylated lysosomal hydrolases						
JOURNAL	Patent: US 6670165-A 21 30-DEC-2003;						
FEATURES	Location/Qualifiers						
source	1..3621						
	/organism="unknown"						
	/mol_type="genomic DNA"						
ORIGIN							
Query Match	62.3%; Score 3485.4; DB 6; Length 3621;						
Best Local Similarity	99.7%; Pred. No. 0;						
Matches 3492; Conservative	0; Mismatches 11; Indels 0; Gaps 0;						
Qy	297	AGCCGAGATCAATACCATGTTTGTTCATTCCTATAGACAATATGCTGGAAGTCC	356				



Db 1195 CTAATGATGATGTCATGTTGGGAAGGATGCTGCGCAGATGATTTTTTACAGTCACTCC 1254  
Qy 1437 AAAGCCAGAGGTTTATTTGACATGGCTGTGCGAACTGTGCGAGGGTCCCGAGGT 1496  
Db 1255 AAAGCCAGAGGTTTATTTGACATGGCTGTGCGAACTGTGCGAGGGTCCCGAGGT 1314  
Qy 1497 TCCTGGATTAGGATGGCTATTGTGACAAAGCTTGTATAATTCAGCCTGGGATGGGAT 1556  
Db 1315 TCCTGGATTAGGATGGCTATTGTGACAAAGCTTGTATAATTCAGCCTGGGATGGGAT 1374  
Qy 1557 GGTGGGATTTGCTCTGGAAAACAGTGGAGGAGTGCCTATATTTGAGGAGGTGGAGTACT 1616  
Db 1375 GGTGGGATTTGCTCTGGAAAACAGTGGAGGAGTGCCTATATTTGAGGAGGTGGAGTACT 1434  
Qy 1617 GGGAGTATTGGATTGGACACCCCTGGCAGTTTGGTGAGGAATAAACAAGTGTCTCTTAC 1676  
Db 1435 GGGAGTATTGGATTGGACACCCCTGGCAGTTTGGTGAGGAATAAACAAGTGTCTCTTAC 1494  
Qy 1677 TGTAATCAGGATTTGCGAATTTCTTGCTGCTGATAGTTTCTGTGACCAAGCATGCAAT 1736  
Db 1495 TGTAATCAGGATTTGCGAATTTCTTGCTGCTGATAGTTTCTGTGACCAAGCATGCAAT 1554  
Qy 1737 GTCCTGCTGCTGGGTTTGTGCTGGCAGTGTGGGCAAGATCATTTTCATGAATTTGAT 1796  
Db 1555 GTCCTGCTGCTGGGTTTGTGCTGGCAGTGTGGGCAAGATCATTTTCATGAATTTGAT 1614  
Qy 1797 AAAGTGATCCTTCTCCCAACACAGACTCACTATATTTATTCAAAAGGTGAATGCTGCT 1856  
Db 1615 AAAGTGATCCTTCTCCCAACACAGACTCACTATATTTATTCAAAAGGTGAATGCTGCT 1674  
Qy 1857 TATTTACGCTTTGAGAGTATGACAAAGAGAGTTGAAGTGGCTATAGTGACAAATCCA 1916  
Db 1675 TATTTACGCTTTGAGAGTATGACAAAGAGAGTTGAAGTGGCTATAGTGACAAATCCA 1734  
Qy 1917 ATAATTCAGACTGCTTCTATTTGCCAAAGTGGAAACCATCTCACTCAATATGCAAGT 1976  
Db 1735 ATAATTCAGACTGCTTCTATTTGCCAAAGTGGAAACCATCTCACTCAATATGCAAGT 1794  
Qy 1977 GGAATGAATGCCAACCAATATTTAATCTCAGTTTCAAAATACAAACGATGAAGAG 2036  
Db 1795 GGAATGAATGCCAACCAATATTTAATCTCAGTTTCAAAATACAAACGATGAAGAG 1854  
Qy 2037 TTCAAAATTCAGATTAACAGTGGAGTGACACAGAGGAGGACCAAACTGAAATCTACG 2096  
Db 1855 TTCAAAATTCAGATTAACAGTGGAGTGACACAGAGGAGGACCAAACTGAAATCTACG 1914  
Qy 2097 GCCAGAGAGGTTACGAAAATTTAGTTAGTCCCATACACTTCTCCAGAGGCGGAAATC 2156  
Db 1915 GCCAGAGAGGTTACGAAAATTTAGTTAGTCCCATACACTTCTCCAGAGGCGGAAATC 1974  
Qy 2157 CTTTTTGAGGATATTTCCCAAGAAAACGCTTCCGAAAGTTTAAAGACATGATTTAAC 2216  
Db 1975 CTTTTTGAGGATATTTCCCAAGAAAACGCTTCCGAAAGTTTAAAGACATGATTTAAC 2034  
Qy 2217 TCAACAAGGAGGCGGAGAGAGGTGAAATTTCCCTGGTAAATATTTTCACTCCTTCCA 2276  
Db 2035 TCAACAAGGAGGCGGAGAGAGGTGAAATTTCCCTGGTAAATATTTTCACTCCTTCCA 2094  
Qy 2277 AAAGACGCGCAGTTGAGTCTCAATACCTTGGATTGCAATTTGCAACATGGAGACATCACT 2336  
Db 2095 AAAGACGCGCAGTTGAGTCTCAATACCTTGGATTGCAATTTGCAACATGGAGACATCACT 2154  
Qy 2337 TTGAAAGGATCAATTTGTCGAAGTCAGCCTTGTGAGATCATTTCTGATCAATCTCAG 2396  
Db 2155 TTGAAAGGATCAATTTGTCGAAGTCAGCCTTGTGAGATCATTTCTGATCAATCTCAG 2214  
Qy 2397 CATGCTAAAATAAAAATCAAGCTATAATAACAGATGAACAAATGACAGTTTGGTGGCT 2456  
Db 2215 CATGCTAAAATAAAAATCAAGCTATAATAACAGATGAACAAATGACAGTTTGGTGGCT 2274  
Qy 2457 CCACAGGAAAACAGGTTTCAAAAAGCATCTTGCCAAAACAGCTTAGGAGTGTCTGAAAGA 2516  
Db 2275 CCACAGGAAAACAGGTTTCAAAAAGCATCTTGCCAAAACAGCTTAGGAGTGTCTGAAAGA 2334

Qy 2517 TTGCAGAGGTTGACTTTTCTCAGTGAGTGATAAAGTGAATGGTTCATGACAGGGTCTAG 2576  
Db 2335 TTGCAGAGGTTGACTTTTCTCAGTGAGTGATAAAGTGAATGGTTCATGACAGGGTCTAG 2394  
Qy 2577 AATCCACCCCTGAGCTTTGGAGACCAAGCAAGATTTAGAGTGGAAACTCACACCCCAAAA 2636  
Db 2395 AATCCACCCCTGAGCTTTGGAGACCAAGCAAGATTTAGAGTGGAAACTCACACCCCAAAA 2454  
Qy 2637 ACCATAGGGGAATGTGCAAAAGAAAAGCCCCATCTCTGATTTGTTCCACTGGAAAGC 2696  
Db 2455 ACCATAGGGGAATGTGCAAAAGAAAAGCCCCATCTCTGATTTGTTCCACTGGAAAGC 2514  
Qy 2697 CAGATCAGAAAAGAAAAGAAAATCAAGGAAAAGAAAAGACAGTAGAATGGAGGAA 2756  
Db 2515 CAGATCAGAAAAGAAAAGAAAATCAAGGAAAAGAAAAGACAGTAGAATGGAGGAA 2574  
Qy 2757 AATGCTGAAAATTCATATGAGGCTTACTGAAAGTGTACTTGGGAAAAGCTGACGATTTAC 2816  
Db 2575 AATGCTGAAAATTCATATGAGGCTTACTGAAAGTGTACTTGGGAAAAGCTGACGATTTAC 2634  
Qy 2817 ACAGATAGTTACTTTGGGCTTTTGGCATGGGAAAAGAAAAGTATTTCCAAAGATCTCTC 2876  
Db 2635 ACAGATAGTTACTTTGGGCTTTTGGCATGGGAAAAGAAAAGTATTTCTAGATCTCTC 2694  
Qy 2877 GACGAAAAGAGTCAATTGGAAGACAAATTTGGCATCTTCACTGATAGCAAAAATACTGGG 2936  
Db 2695 GACGAAAAGAGTCAATTGGAAGACAAATTTGGCATCTTCACTGATAGCAAAAATACTGGG 2754  
Qy 2937 AGGCAACTAAAAGATACATTTGCAGATTTCCCTCAGATATGTAATAAATAATTTCTAATAGC 2996  
Db 2755 AGGCAACTAAAAGATACATTTGCAGATTTCCCTCAGATATGTAATAAATAATTTCTAATAGC 2814  
Qy 2997 AAGTTTGGATTTCATATCGGGAAGTCCCTGCTCAGATGCTCAGATGATGACCGGAT 3056  
Db 2815 AAGTTTGGATTTCATATCGGGAAGTCCCTGCTCAGATGCTCAGATGATGACCGGAT 2874  
Qy 3057 GTTATCGAAGACTGCAAGATATTTCCCTGGAAGATTTTGAACAGCGTCATTTACAAA 3116  
Db 2875 GTTATCGAAGACTGCAAGATATTTCCCTGGAAGATTTTGAACAGCGTCATTTACAAA 2934  
Qy 3117 GTCCGCAATTTGAGGATATGAGTTTGGCTTCTCTTATTTTATTTATCTCATGAGTGA 3176  
Db 2935 GTCCGCAATTTGAGGATATGAGTTTGGCTTCTCTTATTTTATTTATCTCATGAGTGA 2994  
Qy 3177 GTCCGCAATTTGAGGATATGAGTTTGGCTTCTCTTATTTTATTTATCTCATGAGTGA 3236  
Db 2995 GTCCGCAATTTGAGGATATGAGTTTGGCTTCTCTTATTTTATTTATCTCATGAGTGA 3054  
Qy 3237 TTGCTGACAGAGAAATCCGACACTGGCTACCGAATTTACGAACTGCGTTAAGTTTG 3296  
Db 3055 TTGCTGACAGAGAAATCCGACACTGGCTACCGAATTTACGAACTGCGTTAAGTTTG 3114  
Qy 3297 CAGGATTTGACAGGCTGGAACACATGCTTAATAATTTGCTCAAAAATGCTTCTGCTGAT 3356  
Db 3115 CAGGATTTGACAGGCTGGAACACATGCTTAATAATTTGCTCAAAAATGCTTCTGCTGAT 3174  
Qy 3357 ATCAGCAGCTAAATAATTTCCACCACTCAGGATCTCTATGATCCCACTGCCA 3416  
Db 3175 ATCAGCAGCTAAATAATTTCCACCACTCAGGATCTCTATGATCCCACTGCCA 3234  
Qy 3417 CCGGTCACTAAAAGTCTAGTAACAAACTGTAAACCAAGTAACTGACAAAATCCCAAGCA 3476  
Db 3235 CCGGTCACTAAAAGTCTAGTAACAAACTGTAAACCAAGTAACTGACAAAATCCCAAGCA 3294  
Qy 3477 TATAAGGACAAAACAAATATAGGTTTGAATTCATGGGAGAGAAATCGCTTTTAA 3536  
Db 3295 TATAAGGACAAAACAAATATAGGTTTGAATTCATGGGAGAGAAATCGCTTTTAA 3354  
Qy 3537 ATGATTCGTACCAAGCTTCTCATGTTTGGCAGTTGGATGACATAAGAAAACCCCT 3596  
Db 3355 ATGATTCGTACCAAGCTTCTCATGTTTGGCAGTTGGATGACATAAGAAAACCCCT 3414







QY 1059 CCTGCATATTTATTATGGATCTGAGCGCATCAGCGAGTCTAAGCAGGATGAAGATC 1118  
DB 902 CTTGCGTATCTGCTGTGGAGCTGAGTGCCTCAGCCAGTCCAAAGCAGGATGAAGATC 961  
QY 1119 TCTGCGAGTCTGTTTGAAGATAACGAAGAATCTGAGGTACTCAATGCGATCTATCGAGAG 1178  
DB 962 TCTGCGAGCGCTTTGAGGATAATGAAGAGCTGAGGTACTCGCTCGATCTATCGAGAGA 1021  
QY 1179 CATGCACATGGGTTGCGAATAATTTTCAATGTCACCAACGGGCGAGATTCATCTCGGCTG 1238  
DB 1022 CACGCGCATGGGTACGGATAATTTTCAATGTCACCAACGGGCGAGATTCATCTCGGCTG 1081  
QY 1239 AACCTTGAACAATCTCTGAGTGAACAATAGTAACACACACAGGATGTTTTCGAAATTTGAGC 1298  
DB 1082 AACCTTGAACAACCTCTGAGTGACCAATAGTGACCCACACAGGACATTTTCCAAATCTGAGC 1141  
QY 1299 CACTTGCCTACCTTTAGTTACCTGCTAATGAAGTCACTTATCGCATCGAAGGGCTG 1358  
DB 1142 CACTTGCCTACTTTTCAGTTCCCTGCTATTGAAGTCACTTACCCGATCGAAGGGCTG 1201  
QY 1359 TCCCAAGAGTTTATTTACTTAATCATGATGTCATGTTTGGAAAGGATGTCGCGCAGAT 1418  
DB 1202 TCCCAAGAGTTTATTTATTAATGACGATGTCATGTTTCGGTGAAGGCTCTGCGCGAC 1261  
QY 1419 GATTTTACAGTCACTCCAAAGGCCAGAGGTTTATTTGACATGGCCTGTGCCAACTGT 1478  
DB 1262 GATTTTACAGCACTCCAAAGGTCAAAAGGTTTATTTGACATGGCCTGTGCCAACTGT 1321  
QY 1479 GCCAGGGCTGCCAGGTTCTGGAATTAAGATGCTATTGTGACAGGCTTGAATAAT 1538  
DB 1322 GCAGAGGCTGCCCGGCTCTCGGATTAAGACGGCTATTGTGTAAGGCTGTGAATAAC 1381  
QY 1539 TCAGCTCGGATTTGGGATGTTGGGATGCTCTGGAACAGTGGAGGGAGTCCGTATATT 1598  
DB 1382 TCAGCTGTGACTGGGATGGCGGAGCTGCTCTGTTATATCTGAGGGAACGGGTTGTT 1441  
QY 1599 GCAGAGGTGGAGTACTGGGAGTATGGAGTTGGACACCCCTGGCAGTTTGGTGAGGA 1658  
DB 1442 GCAGAGGTGGGGTACCGGGAATTTGGAGTGGACAGCACTGGCAGTTTGGTGAGGA 1501  
QY 1659 ATAAACAGTGTCTTACTGTAAATCAGGAGTGGGAATCTCTGGCTCGCTGATTAAGTTC 1718  
DB 1502 ATAAACACCATCTCTTACTGTAAACAGGATGTGCAAACTCTGGCTGGCTGACAACTTC 1561  
QY 1719 TGTGACCAAGCATCAATGTCTTGTCTGTGGGTTTGTGCTGGCAGTGTGGCAAGAT 1778  
DB 1562 TGTGACCAAGCTGTAACTCTTATCTCTGGGTTTGTGCTGGTGTGCTGAGTGGCAAGAT 1621  
QY 1779 GATTTTCAATGATTAAGTGTATCTCTCCCAACCCAGACTCACTATATTATTTCCA 1838  
DB 1622 CATTTTCATGAATGTATAAGTAACACTTCTCCCAACCCAGACTCACTATGTTGTCCTC 1681  
QY 1839 AAAGGTGAATGCTGCTTATTTCAAGTTTGCAGAAAGTAGCCAAAGAGGAGTTGAAGGT 1898  
DB 1682 AAAGGTGAATACCTGCTTATTTCAAGTTTGCACATAGCCAGAGAGGAGTTGAAGG 1741  
QY 1899 GCTATAGTGAACAATCCCAATAATTCGACATGCTTCTTATTCGCAACAGTGGAAACCATC 1958  
DB 1742 ACCTACAGCGACAACCCCATCATCGCCACGCTCCATTGCAAAACAGTGGAAACCCATA 1801  
QY 1959 CACCTCATATGCAAGTGAATGAATGCCACCAATATACATTTTAACTCAGCTTTTCAA 2018  
DB 1802 CACCTGATATGCAAGTGGATGAACGCAACCCAGATCTATTTTAACTCAGCTTTCAA 1861  
QY 2019 AATACAAACGATGAAGAGTTCAAAATGACAGTGAAGTGGAGGTGGACACAGGAGGGA 2078  
DB 1862 AACGCCAACGACGAAGAGTTCAAGTCCAGATAGCAGTAGAGTGGACACAGGAGGCG 1921  
QY 2079 CCAAACTGAATCTCAGCCCGCAGAGGTTACGAAATTTAGTTAGTCCCATTAACCTT 2138  
DB 1922 CCCAACTGAATTTTACAAACCCAGAGGCTATGAAGTTTGGTTAGCCCACTGACACT 1981  
QY 2139 CTTCCAGAGCGGAAATCTTTTTCAGGATATTCCCAAGAAACCGCTTCCCGAAGTTT 2198

DB 1982 CTTCTCAGGCTGAGCTCCCTTTTGAGATGTCCTCCCAAGAGAAACGCTTCCCAAGATC 2041  
QY 2199 AAGAGACATGATTTAACTCAACAGAGAGGCCAGGAGAGGTGAAAATTTCCCTGGTA 2258  
DB 2042 AGGAGACATGATTAATGCAACAGGAGATTTCAAGAGAGGTGAAAATTTCCCGGGTA 2101  
QY 2259 AATATTTCTCTCTCCAAAGAGCGCCAGTGTGAGTCTCAATACCTTTGATTTGCAACTG 2318  
DB 2102 AATATTTCTCTCTCCAAAGAGGCCAGTGTGAGTGTGAGCACTTTGATTTGCAACTA 2161  
QY 2319 GAACATGGAGACATCACTTTTGAAGGATCAAAATTTGTCAAAGTCAAGCTTGTGTGATCA 2378  
DB 2162 GAACGTGGAGACATCACTCTGAAAGGATATACTTGTCCAAGTCAAGCTTGTGTGATCT 2221  
QY 2379 TTTCTGATGAACATCAAGCATGCTATAATAAATAAATAAATAAATAAATAAATAAATAA 2438  
DB 2222 TTTCTGGGAAATTCATAGATATAAATAAATAAATAAATAAATAAATAAATAAATAA 2278  
QY 2439 AATGACAGTTTGGTGGCTCCACAGGAAACAGGTTTATAAAGCATCTTGTCCAAACAGC 2498  
DB 2279 AAAGGCAACCTGAGGTCCACAGGAAACCCCTTCTCACAGAG-----TCCACATGGC 2332  
QY 2499 TTAGGAGTGTCTGAAAGATTGACAGGTTGACTTTTCTTCTGAGTGTGATAAGTGAAT 2558  
DB 2333 TTTGCTGGTGAACACAGATCAGAGATGGACTGCCCCAGCAGAGAGTGAACCGTGA 2392  
QY 2559 GGTGATGACAGGTCAGAAATCCACCCCTGGACTTCGAGACACACAGAGATTTAGAGTG 2618  
DB 2393 GGCGGTGACCAACCTTTGATTCACCCCGGTTTGGAGCAATGCAAGATT-----2445  
QY 2619 GAAACTCACACCCCAAAAAACCATAGCGGAAATGTGACAAAAAGAGAGCCCATCTCTG 2678  
DB 2446 -----GGCCCGACCTACACTAGCGTGACTGTGTCCAAAGAGAACTTTTCCCGCTG 2497  
QY 2679 ATTGTTCACATGGAAGCCAGATGACAAAGAAAGAAATACACAGGGAAGAAAGAG 2738  
DB 2498 ATCTTTCCCGCAAGAACCCACTTGCAGAAAGAGAGAGTGAACGGGAGAG-----2551  
QY 2739 AACAGTAGAATGGAGAAATGCTGAAATCACATAGGCGTTACTCAAGTGTACTTGA 2798  
DB 2552 -----GGCAATGCTGTACTGTAAAGAGTAGTGCCTGGC 2587  
QY 2799 AGAAGCTGAGCATTTACAGATGTTACTTGGGCTTTTGGCCATGGGAGAAAAAG 2858  
DB 2588 AGACGTTGAGC-----AGAAATTTCCAGGCTTTTGGCCTGGGAGAAAAAG 2638  
QY 2859 TATTTCCAGATCTTCTCGACGAAGAGAGTCAATGAAGACACAAATTTGGCATACTTCACT 2918  
DB 2639 TATTTCCAGACCTTCTTGTGAGGAGAGTCAATGAAGACCCAGTGGGTACTTTACA 2698  
QY 2919 GATAGCAAAATATCTGGAGGCAACTAAAAGATACATTTGACAGATTCCTTCAGATATGA 2978  
DB 2699 GACAGCAAAATATCTGGAGGCAACTAAAAGATACATTTGACAGCTCCCTCCGATACGTC 2758  
QY 2979 AATTAATTTCTAATAGCAAGTTTGGATTCACATCGGAAAGTCCCTGCTCACATGCCCT 3038  
DB 2759 AATTAATTTCTAAGCAAGTTTGGATTCACATCGGAAAGTCCCTGACACATGCCCT 2818  
QY 3039 CACATGATTCACCGGATTTGTTATGCAAGAACTGCAAGATATGTTCCCTGAGAAATTTGAC 3098  
DB 2819 CACATGATTCACAGGATCGTTATGCAAGAACTCCAAGATATGTTCCCTGAGAAATTTGAC 2878  
QY 3099 AGACGTCATTTCAAAAGTGGCCATTTCTGAGGATATGAGGTTTGGCTTCTTATTTT 3158  
DB 2879 AAGACTTCATTTCAAGGTCGCTCTCTGAGGACATGCAAGTTTGGCTTCTCTACTTT 2938  
QY 3159 TATTTCTCATGAGTGCAGTGCAGCACTGAAATATATCTCAAGTCTTTGATGAAGTTGAT 3218  
DB 2939 TATTTCTCATGAGTGCAGTGCAGTGCAGCTTCAAGCTTATTTCCCAAGTCTTTTCAAGTAGAC 2998  
QY 3219 ACAGATCAATCTCGTGTCTGTCTGACAGAGAAATCCGAACACTTGGTCCAGAAATTCAC 3278



Db 2999 ACAGACCAATCTGGTGTCTTCTGTGATAGGGAATCCGAACACTGGCCACGAGAAATTCAC 3058  
QY 3279 GAACTGCGTTAAGTTTCAGAGATTGACAGGTCTGGAACACATGCTAATAAATGCTCA 3338  
Db 3059 GACCTACCTTTAAGCTTCAGAGATTGACAGGTTCGGAACACATGTTAATAAATGCTCA 3118  
QY 3339 AAAATGCTTCCTGCTGATATCAGCAGCTTAATTAATAATCCACCACTCAGGATCCCTAC 3398  
Db 3119 AAAATGCTTCCTGCTGATATCAGCAGCTTAATTAATAATCCACCACTCAGGATCCCTAC 3178  
QY 3399 TATGATCCCAACCTGCCCGCTCTCTAATAAGTCTAGTAAACAACTGTAAACCACTAGTAACT 3458  
Db 3179 TACGACCCCAACCTGCCCGCTCTCTAATAAGTCTAGTAAACAACTGTAAACCACTAGTAACT 3238  
QY 3459 GACAAATCCCAAGGATATAGGACAAACAAATATAGGTTTGAATCATGGGAGAA 3518  
Db 3239 GACAAATCCCAAGGATATAGGACAAACAAATATAGGTTTGAATCATGGGAGAG 3298  
QY 3519 GAAAGAAATCGCTTTTAAATGATTCGTACCAACGTTTCTCATGTGTTGGCCAGTTGGAT 3578  
Db 3299 GAAGAAATCGCTTTTAAATGATTCGTACCAACGTTTCTCATGTGTTGGCCAGTTGGAT 3358  
QY 3579 GACATAAGAAAACCTTAGGAAGTTTGTTCCTGCTGAATGACAACTGACCAATCAT 3638  
Db 3359 GACATCAGAAAACCTTAGGAAGTTTGTTCCTGCTGAATGACAACTGACCAATCAT 3418  
QY 3639 AAAGATGCTCAGACAGTCAAGGCTGTTCTCAGGACCTCTATGATTAATCATGTTCCCCATA 3698  
Db 3419 AAAGATGCTCAGACAGTCAAGGCTGTTCTCAGGACCTCTATGATTAATCATGTTCCCCATA 3478  
QY 3699 CTTTCCCAATTGAACTGCCAAGAGATCGAAACCGTTTCTTCATATGATGATGAGCTG 3758  
Db 3479 CTTTCCCAATTGAGCTGCCAAGAGATCGAAACCGTTTCTTCATATGATGATGAGCTG 3538  
QY 3759 CAGGAATCGGGCTTATCGAGACAAATGGAATTTGGACCCATCTGTCTAGCAACA 3818  
Db 3539 CAAGAAATGGGGGCTTATCGAGACAAATGGAATTTGGACCCATCTGTCTAGCAACA 3598  
QY 3819 TTGATTATGTTTACTATATCTCATATTTTCTGCTGAGCAGTTAATTTGCACTTAAGCGAAG 3878  
Db 3599 TTGATTATGTTTACTATATCTCATATTTTCTGCTGAGCAGTTAATTTGCACTTAAGCGAAG 3658  
QY 3879 ATATTTCCCAAGAGAGATACAAAGAACTAGTCCCAATCGAATCAGAGTATAGAAG 3938  
Db 3659 ATATTTCCCAAGAGAGATACAAAGAACTAGTCCCAATCGAATCAGAGTATAGAAG 3718  
QY 3939 ATCTTCATTTGAAAACCTTACTCTCAGCATTACTGAGCATTATTAAGCTTCAGCTTCAC 3998  
Db 3719 ATCTTCATTTGAAAACCTTACTCTCAGCATTACTGAGCATTATTAAGCTTCAGCTTCAC 3777  
QY 3999 AGAGATGCTTTGTGATGTGATGCTTA-GCAGTTTGGCCCAAGAGAAATATCCAGT 4057  
Db 3778 AGCGAGTCTCTGTGATGTGACAGAGGCGACCTCGTGGGAGAGAGGACATCGTGACG 3837  
QY 4058 ACCATGCTGTTTGTGGCATGATATATGACCACTGACTAGGAATTTATTAACCAACCCAC 4117  
Db 3838 ACCGGGTTCTTCTGCAATGGGAAGAGAGCCCACTGACTGGAATTTATTTTCCACAC 3893  
QY 4118 TGAATACTGTGTGTCAGAGCTCTGAACTGATTTTACTTTTAAAGAAATTTGCTCATGG 4177  
Db 3894 TAAGAACTGTGTCAAATAGC---TTGTACAGCTTGTCTTTTAAAGAAATTTGCGGAAGG 3949  
QY 4178 ACCGTGCTATCTTTTATTAAGAAAGGCTCAGTCAAGAGAGAGCTGTATTTTCCACAG 4237  
Db 3950 ACCGTGCGCTGTGTGACAAACCCCTCCCTGACA-----AGCTGTGTTTCTTCCCC 4001  
QY 4238 CAATCATTTGACAGCTTAATTTATAGGAGAGCCCTATGCCAGCTGGGAGTGAATGCTAAG 4297  
Db 4002 CAGTTACTGCAGACT-----GAGAAACAGTCCATCTTGAAGCAAGTGGGAG 4050  
QY 4298 AGGCTCAGCTCTTTCATTTCAAGGCTTTTGTCTTAAAGTTTTCAGCTTTT-----TTT 4350  
Db 4051 GGGCCCCAGCTTTTGTGCTTCAAGAGCTTTTCAGCATAATTTCTGGCTGTCTCTCTCTCT 4110

QY 4351 TTTTCATTTCCCATTTT-----AAGTAGTTACTAAGTTAACTAGTTATTTCTTGC 4400  
Db 4111 GATCCATTTCCCATTTT-----TTTTAAAAACCAATAAGTGGCTACTAAGTTAGTCACTCAC 4170  
QY 4401 TTCTGAGTATAACGAATTTGGGATGTCTAAACCTATTTTATAGATGTTTATTTAAATAATG 4460  
Db 4171 TTCTCAAATAACAATCAGGATGTCAAAAC---ATTGTATAGATCTTATTTAAATAATA 4228  
QY 4461 CAGCAATATCAGCTCTTATTTAGCAATACCTAAATATGAGTTTATTTAATTTAAGACT 4520  
Db 4229 TAGAAGGATTTACTTCTTT---AGCCTATCTAAATTTATTTTATTTAAACAGTCAAGTG 4285  
QY 4521 GTAAATGGTCTTTAAACCACTAATCTGAAGAGCTCAATGATTGACATCTGAAATGCTTT 4580  
Db 4286 GTCTTGAATCGCTAAC-----AATCTAGTGAAGCTCGA-GATTGACGTTGAAGTGTCTT 4340  
QY 4581 GTAAATTTAGCTTCAGCCCTTAAGAACTGCTATGATTTCAAGTGCAGTCTTAATTTCAAC 4640  
Db 4341 GAGCTTGTTTAACTCATTTCCCAAGATACTGTGACCTCGTGTGGGGCTGATTGCGAA 4400  
QY 4641 AGCTAGAGT-----TAGTACTACTTACCAGATGTAATATGTTTGGAAATGTACAT 4693  
Db 4401 GGGCTAGTGTACAGTACGAGTGTCTCTCAGGATGTAATATGCTGTGGAAATGTACAT 4460  
QY 4694 ATTCAAAACAGAAAGTGCCTCAITTTTAAAGAAATGAGTAGTGTCTGATGGCACTGGCACATTACA 4753  
Db 4461 A--CAACAAGAGTGCCTCAGTGTCTCAGGATGTAATATGCTGTGGAAATGTACAT 4518  
QY 4754 GTG 4756  
Db 4519 GTG 4521

RESULT 15  
AR300468  
LOCUS AR300468 5229 bp DNA linear PAT 12-JUN-2003  
DEFINITION Sequence 16 from patent US 6537785.  
ACCESSION AR300468  
VERSION AR300468.1 GI:31687910  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 5229)  
AUTHORS Canfield, W.M.  
TITLE Methods of treating lysosomal storage diseases  
JOURNAL Patent: US 6537785-A 16 25-MAR-2003;  
FEATURES  
source  
1..5229  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN

Query Match 50.4%; Score 2820.8; DB 6; Length 5229;  
Best Local Similarity 79.1%; Pred. No. 0;  
Matches 3667; Conservative 1; Mismatches 833; Indels 137; Gaps 21;  
QY 150 GSCGCTGAAGGGGTGATGCTGTTCAAGCTCTCTGAGAGACAAACCTATACCTGCTGCTGCC 209  
Db 1 GSCGCTGAAGGGGTGATGCTGTTCAAGCTCTCTGAGAGACAGACCTATACCTGCTGCTGCC 60  
QY 210 CACAGGTATGGCTCTAGCTGTGCTTCTTGGCGCTGTTGTCACCATGCTTCCGCCCTTC 269  
Db 61 CACAGGTATGGCTCTAGCTGTGCTTCTTGGCGCTGTTGTCACCATGCTTCTGCTTTC 120  
QY 270 CAGTTCCGAGAGGTGGTCTTCTGGAATGGAGCCGAGATCAATACCATGTTTCTTGTGATTC 329  
Db 121 CAGTTCCGAGAGGTGGTCTTCTGGAATGGAGCCGAGATCAGTACCATGTTTCTTGTGATTC 180  
QY 330 TATAGACAAATATGCTGGAAGTCTTTTTCAGAAATCGGCTTTGCTGCCCATGCCGATT 389  
Db 181 TATAGACAAATATGCTGGAAGTCTTTTTCAGAAATCGGCTTTGCTGCCCATGCCGATT 240



Db 2392 GTGACCGTGAAGGCGGTGACCAAGCTTTGAAATCCACCCCGGTTGGAGACCAATGCA 2451  
Qy 2607 AGATTAGAGTGGAACCTCACACCCAAACAAACCTAGGCGGAATGTGACAAAAGAAAAG 2666  
Db 2452 AGATT-----GGCCCGCCTACACTAGCGGTGACTGTGTCCAAAGAAAC 2496  
Qy 2667 CCCCACATCTCTGATTGTTCCACTGGAAGGCGAGATGACAAAGAAAAGAAAATCACAGG 2726  
Db 2497 CTTTCACCGCTGATCGTTCCCGGAGAACCACTTGCCTAAAGAGAGGAGGTGACAGG 2556  
Qy 2727 AAAGAAAAGAGAACAGTAGAATGGAGAAAATGCTGAAAATCAATAGCGGTTACTGAA 2786  
Db 2557 GCAGAA-----GGCAATGCTGTACTGTAAAGGAG 2586  
Qy 2787 GTGTTACTTTGGAGAAAGCTGCAGCAATTACACAGATAGTTACTTTGGGCTTTTGGCCATGG 2846  
Db 2587 TTAGTGCTTGGCAGACGCTTGAGC-----AGAAATATCCAGGCTTTTGGCCCTGG 2637  
Qy 2847 GAGAAAAAAGTATTTCCAAAGATCTTCTCGACGAAGAAGAGTCAITTAAGACACAAATTG 2906  
Db 2638 GAGAAAAAAGTATTTCCAAAGCTTCTTGATGAGGAAGAGTCAITTAAGAGCCAGTTG 2697  
Qy 2907 GCATACTTTCATGATAGCAAAAATCTGGGAGGCACTAAAGATACATTTGCGAGATCC 2966  
Db 2698 GCGTACTTTACAGACCGCAACATACCGGAGGCACTTAAAGATACATTTGCGAGACTCC 2757  
Qy 2967 CTCAGATATGAAAAATTTCTAAATAGCAAGTTTGGATTCACATCGCGGAAAGTCCCT 3026  
Db 2758 CTCAGTACGTCATTAATTTCTCAACAGCAAGTTTGGATTCACATCCAGGAAGTCCCT 2817  
Qy 3027 GCTCATGCTTCATGATGACCGGATTTGATGCAAGAACTGCAAGATATGTTCCCT 3086  
Db 2818 GCACACATGCGGCACATGATTGACAGGATCGTTATGCAAGAACTCCAAGATATGTTCCCT 2877  
Qy 3087 GAAGAAATTTGACAGACGTCATTTACAAAGTGCGCCACTCTGAGGATATGSCAGTTTGC 3146  
Db 2878 GAAGAAATTTGACAGACTTCATTTCAAGAGTGCGTCACTCTGAGGACATGCGAGTTTGC 2937  
Qy 3147 TTTCTCTATTTTATATCTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTTT 3206  
Db 2938 TTTCTCTACTTTTATCTCATGATGATGAGTGCAGTTCAGCCCTCAATATTTCCCAAGTCTTT 2997  
Qy 3207 GATGAAGTTGATACAGATCAATCTGGTCTGTGTGTCAGAGAAATCCGACACTGGCT 3266  
Db 2998 CATGAAGTAGACACAGACCAATCTGTGTCTGTCTGATAGGAAATCCGAAACCTGCGCC 3057  
Qy 3267 ACCAGAAATTCAGAACTCCGTTAAAGTTTGAGGATTTGACAGGCTCTGGAACACATGCTA 3326  
Db 3058 ACGAGAAATTCAGACCTACTTTAAGCTTGAGGATTTGACAGGTTTGGAAACACATGTTA 3117  
Qy 3327 ATAAATTTGCTCAAAATGCTTCTGCTGATATACAGGAGTAAATATATTTCCACCACT 3386  
Db 3118 ATAAATTTGCTCAAAATGCTTCCCGCTAATATATCACTCAACTCAACATCCCAACCGACT 3177  
Qy 3387 CAGGAATCTCTATGATCCCAACCTGCGACCGGTGACTAAAGTCTAGTAACCAAACTGT 3446  
Db 3178 CAGGAGACATACAGACCCCAACCTGCTCGGTCACTAGAGTCTTTGTCACCAACTGT 3237  
Qy 3447 AAACAGTAATGCAAAATTCACAAAGCATATTAAGGCAAAAAACAAATATAGGTTTGA 3506  
Db 3238 AAGCCAGTAATGCAAGATCCAAAGCTTATAAGACAAAGAAACAAATACAGGTTTGA 3297  
Qy 3507 ATCATGGGAGAGAGAAATCGCTTTTAAATGATTCGTACCAACGTTTCTCATGTGTT 3566  
Db 3298 ATCATGGGAGAGAGAAATCGCTTTCAAGATGATACGAACCAATGTTTCTCATGTGTT 3357  
Qy 3567 GCGCAGTTGGATGACATAGAAAAAACCTTAGGAAGTTTGTTCCTCGAATGACAACT 3626  
Db 3358 GGTCAAGTTGGATGACATCAGAAAAAACCCCAAGGAAGTTGTTGCTCGAATGACAACT 3417  
Qy 3627 GACACATATTAAGTGTCTCAGACGTGAGGCTGTTCTCAGGACTTCTATGATCC 3686

Db 3418 GACCACACCAATAAGATGCCCGGACAGTGAAGGCTGTCTCTCAGGACTTCTATGAGTCC 3477  
Qy 3687 ATGTTCCCATACCTTCCCAATTTGAACTGCAAGAGAGTATCGAAACCGTTTCTCTCAT 3746  
Db 3478 ATGTTTCCCATACCTTCCCAATTTGAGCTGCAAGAGAGTATCGGAACCGTTTCTGCAC 3537  
Qy 3747 ATGCTATGAGCTGCAGGAATGAGGCTTATCGAGACAAATTTGAAGTTTTTGACCACTGT 3806  
Db 3538 ATGCTATGAGCTTCCAAAGATGGCGCATATCGAGACAAAGTGAAGTTTTTGACCACTGC 3597  
Qy 3807 GTACTAGCAACATTTGATTATGTTTACTATTTCTCATTTTGTCTGAGCAGTTAATTGCA 3866  
Db 3598 GTACTAGCAACCTTGATTATATTTACTATTTCTCATTTTGTCTGACACATAATTGCT 3657  
Qy 3867 CTTAAGCGGAAGATATTTCCAGAAAGGAGGATACACAAAGAGCTAGTCCCAATCGAATC 3926  
Db 3658 CTGAAGCGGAAGATATTTCCAGAGGAGGAGTACACAAAGAGCTAGTCCAGCCGAATC 3717  
Qy 3927 AGAGTATAGAGATCTTCAATTTGAAAAACATCTACTCAGCAATTTACTGAGCAATTTAAA 3986  
Db 3718 AGGGTGTAGAGATCTTCAATTTGAAAAAGTCACTACTCTTAGCATCT-GTGAACATCTCCCT 3776  
Qy 3987 ACTCAGCTTCAGAGATGCTTCTGTGATGTGATGCTTA-GCAGTTTGGCCGGAAGAAG 4045  
Db 3777 CTTGACACACACAGCGGAGTCCCTGTGTGGACAGAGCGCTCGTGGGAGAAG 3836  
Qy 4046 AAAATATCCAGTACCATGCTGTTTTGTGCGCATGAATATAGCCACTGACTAGGAATTTAT 4105  
Db 3837 GACATGTCGACACCGGGTCTTCTGCAATGGGAAGAGAGCCCACTGACCTGGAATTTAT 3896  
Qy 4106 TAACCAACCACTGAAACTTGTGTGCGAGCTCTGAACTGATTTTACTTTTAAAGA 4165  
Db 3897 ----CAGCACATAGAACCTGTCAATAGC----TTGTACAGCTTGACTTTTAAAGG 3948  
Qy 4166 ATTTGCTCATGACCTGTCTCATCTTTTATAAAAAGGCTCACTGACAAAGACAGCTGT 4225  
Db 3949 ATTTGCGGAGACCTGTGCGCTTGTGACAAACCTCCCTGACA-----AGCTGCT 4000  
Qy 4226 AATTTCCACACCAATCATGACAGACTAATTTATAGAGAGAGCTATGCCAGCTGGGA 4285  
Db 4001 GGTTCCTTCCCGCAGTACTGCAGACT-----GAGAAACCACTCCATCTTGAAA 4049  
Qy 4286 GTGATTGCTTAAGAGGCTCCAGTCTTGTGATTCCAAAGCTTTTGTAAAGTTTTTGCACTT 4345  
Db 4050 GCAAGTGGGAGGCGGCCAGTCTTGTGATTCCAAAGCTTTCCAGCAATATTTCTGGCTT 4109  
Qy 4346 TT-----TTTTTTTCAATTTCCCAATTTT-----AAGTAGTTACTAAGTTAACT 4388  
Db 4110 GTCTCTCTCTTTGATCCATTTCCCATTTTAAAAAACAATAAGTGGCTACTAAGTT 4169  
Qy 4389 AGTTATTTCTGCTCTGAGTATAAGAAATGGGATGTCTAAACCTATTTTATATAGATGT 4448  
Db 4170 AGTCATTTCTCACTTTTCAAAATPAACAATCAGGATGTCAAAAC--ATTGTATAGATCTT 4227  
Qy 4449 ATTTAAATTAATGACCAATATCACCTCTTATTTGACAAATACCTPAAATATGAGTTTATTA 4508  
Db 4228 ATTTAAATTAATAGAACGATTACTTCT--TTAGCCTATCTAAATTTATTTTATTA 4285  
Qy 4509 ATATTTAAGACTGTAAAGTGTCTTAAACCACT---AAGTCTGAGAGCTCAATGATGA 4565  
Db 4286 ACAGTCAAG-----TGGTCTTGAACCGCTTAACTACTTACAGAGCTCGA-GATTGA 4336  
Qy 4566 CATCTGAAATGCTTTGTATTAATTATGACTTCAGCCCTCAAGATGCTATGATTTACGTGC 4625  
Db 4337 CGTTGAAAGTGTCTTGAGCTTGTTTAACTCATTTCCCAAGAAATACTGTGACCTCTGTGC 4396  
Qy 4626 AGGTCTAATTTCAACAGGCTAGAGT-----TAGTACTATTACAGATGATTAATGT 4678  
Db 4397 GGGCTGTGTTGGAAGGCGCTAGTGTCAAGTGTGCTGCTCAGCGATGTAAATTTATGT 4456  
Qy 4679 TTTGAAATGTACATATTAACACAGAAAGTGTCTCATTTTAAAGATGAGTAGTGTGATGG 4738  
Db 4457 CGTGGAAATGTACATA--CAGACAAAGTGTCTCACTCAGAAATGAGTAGTGTGATGG 4514

Qy 4739 CACTGGCACATTACAGTG 4756  
|||  
Db 4515 CACCAGCGAGTGATGGTG 4532

Search completed: November 22, 2004, 06:14:41  
Job time : 22750 secs

***This Page Blank (uspio)***



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2004, 18:18:51 ; Search time 2385 Seconds  
(without alignments)  
12319.082 Million cell updates/sec

Title: US-10-023-888-3  
Perfect score: 5597  
Sequence: 1 cggagcagggcggtccg.....aaaaagttaatttgaaaa 5597

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues  
Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_23Sep04:.\*  
1: Geneseqn1980s:.\*  
2: Geneseqn1990s:.\*  
3: Geneseqn2000s:.\*  
4: Geneseqn2001as:.\*  
5: Geneseqn2001bs:.\*  
6: Geneseqn2002as:.\*  
7: Geneseqn2002bs:.\*  
8: Geneseqn2003as:.\*  
9: Geneseqn2003bs:.\*  
10: Geneseqn2003cs:.\*  
11: Geneseqn2003ds:.\*  
12: Geneseqn2004s:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5597	100.0	5597	4 ACC81001	Acc81001 Human Glc
2	5597	100.0	5597	10 ADD27811	Add27811 Human Glc
3	5597	100.0	5597	10 RAD62491	Rad62491 Human Glc
4	5597	100.0	5597	10 RAD62650	Rad62650 Human Glc
5	4326	77.3	4369	6 RAD41199	Rad41199 Human nuc
6	4057	72.5	4277	6 RAD41203	Rad41203 Human nuc
7	3761	67.2	3783	4 ACC81010	Acc81010 Human Glc
8	3485.4	62.3	3621	4 ACC81011	Acc81011 Human Glc
9	3465.6	61.9	3600	10 ADD27809	Add27809 Soluble h
10	3465.6	61.9	3600	10 RAD62490	Rad62490 N-acetylgl
11	3465.6	61.9	3600	10 RAD62649	Rad62649 N-acetylgl
12	2820.8	50.4	5229	4 ACC81007	Acc81007 Murine Gl
13	2820.8	50.4	5229	10 ADD27816	Add27816 GlcNAc-ph
14	2820.8	50.4	5229	10 RAD62493	Rad62493 Mouse cDN
15	2820.8	50.4	5229	10 AAD62652	Aad62652 Mouse DNA
16	2613	45.7	8810	10 ADG39793	Adg39793 Human cDN
17	2613	45.7	8811	8 ABX10235	Abx10235 Human cDN
18	2587.4	46.2	2603	10 ADAS3185	Ada53185 Human cod
19	2372.2	42.4	2428	6 AAL50831	Aal50831 Human can
20	2018.2	36.1	2076	4 AAH14758	Aah14758 Human cDN
21	1716.4	30.7	16800	4 AAK82065	Aak82065 Human imm

C	22	1716.4	30.7	16805	4	AAK82067	Aak82067 Human imm
C	23	1716.4	30.7	16806	4	AAK82066	Aak82066 Human imm
C	24	1105.2	19.7	1194	12	ACH89502	Ach89502 Human gen
C	25	1104.8	19.7	1962	4	AAI12977	Aai12977 Probe #29
C	26	1104.8	19.7	1962	4	AAI34336	Aai34336 Probe #30
C	27	1104.8	19.7	1962	4	ABA44229	Abaa44229 Human bre
C	28	1104.8	19.7	1962	4	AAK02966	Aak02966 Human bra
C	29	1104.8	19.7	1962	5	AAI02895	Aai02895 Probe #28
C	30	1104.8	19.7	1962	6	ABS02920	Abs02920 Human gen
C	31	1103.4	19.7	1125	4	AAI22215	Aai22215 Probe #12
C	32	1103.4	19.7	1125	4	AAI47512	Aai47512 Probe #16
C	33	1103.4	19.7	1125	4	ABA49383	Abaa49383 Human bra
C	34	1103.4	19.7	1125	4	AAK15732	Aak15732 Human bra
C	35	1103.4	19.7	1125	5	AAI07914	Aai07914 Probe #79
C	36	1103.4	19.7	1125	6	ABS15475	Abs15475 Human gen
C	37	724.8	12.9	728	10	ADD26805	Add26805 Human adi
C	38	692	12.4	780	4	AAH07289	Aah07289 Human cDN
C	39	501	9.0	526	2	AAV86143	Aav86143 EST clone
C	40	486.2	8.7	830	8	ABX34834	Abx34834 Human mdd
C	41	415.4	7.4	465	9	ACH38230	Ach38230 Human end
C	42	398.4	7.1	537	3	AAA43989	Aaa43989 Mouse sec
C	43	377.2	6.7	459	9	ACH40463	Ach40463 Human foe
C	44	356.2	6.4	392	4	AAH12279	Aah12279 Human cDN
C	45	353.4	6.3	773	10	ADI21775	Adi21775 Novel hum

ALIGNMENTS

RESULT 1  
ACCS81001  
ID ACC81001 standard; cDNA; 5597 BP.  
XX  
AC ACC81001;  
XX  
DT 01-AUG-2003 (first entry)  
XX  
DE Human GlcNAc-phosphotransferase alpha/beta-subunit cDNA.  
XX  
KW Human; N-acetylglucosamine-1-phosphotransferase; nephrotropic;  
KW GlcNAc-phosphotransferase; phosphodiester alpha-GlcNAcase;  
KW N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase;  
KW enzyme replacement therapy; phosphorylated lysosomal hydrolase;  
KW lysosomal storage disease; enzyme; alpha-subunit; gene; ss; beta-subunit.  
XX  
OS Homo sapiens.  
XX  
PH Key Location/Qualifiers  
PT CDS 165...2948  
FT /\*tag= a  
FT /product= "GlcNAc-phosphotransferase alpha subunit"  
FT /note= "No stop codon given"  
FT CDS 2949..3935  
FT /\*tag= b  
FT /product= "GlcNAc-phosphotransferase beta subunit"  
FT /note= "No start codon"  
XX  
US US6537785-B1.  
XX  
PD 25-MAR-2003.  
XX  
PF 10-AUG-2000; 2000US-00636077.  
XX  
PR 14-SEP-1999; 99US-0153831P.  
XX  
PA (GENZ-) GENZYME GLYCOBIOLOGY RES INST INC.  
XX  
PI Canfield WM;  
XX  
DR WPI; 2001-290356/30.  
XX  
DR P-FSDB; ABR61376, ABR61377.  
XX  
PT Novel N-acetylglucosamine-1-phosphotransferase and N-acetylglucosamine-i-



[illegible]



DR WPI; 2003-801323/75.  
DR P-PSDB; ADD27812, ADD27813.  
XX Phosphorylating a protein for treating a patient suffering from a  
PT lysosomal storage disease e.g. Fabry's disease by contacting the protein  
PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated  
XX protein.  
XX  
PS Disclosure; SEQ ID NO 3; 55pp; English.  
XX  
XX The invention relates to a method of phosphorylating a protein comprising  
CC contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-  
CC acetylglucosamine) and producing a phosphorylated protein. the method is  
CC useful for treating a patient suffering from a lysosomal storage disease  
CC e.g. Fabry's disease. The present sequence represents the human GlcNAc-  
CC phosphotransferase alpha/beta subunit precursor cDNA.  
XX  
SQ Sequence 5597 BP; 1678 A; 1151 C; 1219 G; 1549 T; 0 U; 0 Other;  
Query Match 100.0%; Score 5597; DB 10; Length 5597;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CGGAGCCGAGCGGGCGTCCGTCGCGGAGCTGCAATGAGCGCGCGGAGGCTGTGACC 60  
DB 1 CGGAGCCGAGCGGGCGTCCGTCGCGGAGCTGCAATGAGCGCGCGGAGGCTGTGACC 60  
QY 61 TCGCGCGCGCGCCCGACCGGGGCCCTTAATGGCGCTCGCTGAGCGCGCGGCGGCGC 120  
DB 61 TCGCGCGCGCGCCCGACCGGGGCCCTTAATGGCGCTCGCTGAGCGCGCGGCGGCGC 120  
QY 121 GCGGCTCAGGCTCTCGGGCGTGGCGTGGCGGTGAAGGGGTGATGCTGTTCAAGCTCC 180  
DB 121 GCGGCTCAGGCTCTCGGGCGTGGCGTGGCGGTGAAGGGGTGATGCTGTTCAAGCTCC 180  
QY 181 TGCAGAGACAAACCTATACCTGCTGCCAGAGTATGGGCTCTACGCTGTCTTCTGG 240  
DB 181 TGCAGAGACAAACCTATACCTGCTGCCAGAGTATGGGCTCTACGCTGTCTTCTGG 240  
QY 241 GCGTGTGTACCATGCTCCGCTTCGAGTTCGAGAGAGGTGTTCTGGAATGGAGCC 300  
DB 241 GCGTGTGTACCATGCTCCGCTTCGAGTTCGAGAGAGGTGTTCTGGAATGGAGCC 300  
QY 301 GAGATCAATACCATGTTGTTGTTGATTCCTATAGAGACAAATATGCTGGAAGTCCCTTC 360  
DB 301 GAGATCAATACCATGTTGTTGTTGATTCCTATAGAGACAAATATGCTGGAAGTCCCTTC 360  
QY 361 AGAATCGGCTTGTCTGCCCATGCGCATGACGTTGTTTACCTGGGTGAATGGCACAG 420  
DB 361 AGAATCGGCTTGTCTGCCCATGCGCATGACGTTGTTTACCTGGGTGAATGGCACAG 420  
QY 421 ATCTTGAATCTGTAAGGAACCTACAGAGGTACAGAGACAGATGGAGGAGGAGCAAG 480  
DB 421 ATCTTGAATCTGTAAGGAACCTACAGAGGTACAGAGACAGATGGAGGAGGAGCAAG 480  
QY 481 CAATGAGAGAAATCCTTGGGAAAAACACAAACGGAACCTACTAAGAGAGTGAAGCAGT 540  
DB 481 CAATGAGAGAAATCCTTGGGAAAAACACAAACGGAACCTACTAAGAGAGTGAAGCAGT 540  
QY 541 TAGAGTGTGTGTAACACACTGATAGGTGCGCAATGCTTGTACTGACCCAGCCCTGC 600  
DB 541 TAGAGTGTGTGTAACACACTGATAGGTGCGCAATGCTTGTACTGACCCAGCCCTGC 600  
QY 601 CAGCCAAACATCACCCTGAGGACGTCATCTCTTTATCTCTTTCATCTGCGCAGT 660  
DB 601 CAGCCAAACATCACCCTGAGGACGTCGCAATCTCTTTATCTCTTTCATCTGCGCAGT 660  
QY 661 ACATTTCAATGTTGAAAAACAAAAACCCCTTCTACCAATGTCAGTGTGTTGTTTG 720  
DB 661 ACATTTCAATGTTGAAAAACAAAAACCCCTTCTACCAATGTCAGTGTGTTGTTTG 720  
QY 721 ACAGTACTAAGAGTGTGAGATGCCACTCTGAGTCTTAAAGGAATAGCAGACAGA 780

DB 721 ACAGTACTAAGAGTGTGAGATGCCACTCTCTGGAAGTCTTAAAGGAAATAGCAGACAGA 780  
QY 781 CAGTATGAGGGGTACTTGCACAAAGATGAAAGTCTCCTGATAGTCTTAATGCAAG 840  
DB 781 CAGTATGAGGGGTACTTGCACAAAGATGAAAGTCTCCTGATAGTCTTAATGCAAG 840  
QY 841 ATTTGGCTTTCCTGAGTGGATTTCCACCAACATTCAGGAAACAAATCAACTAAAAACAA 900  
DB 841 ATTTGGCTTTCCTGAGTGGATTTCCACCAACATTCAGGAAACAAATCAACTAAAAACAA 900  
QY 901 AATTGCCAGAAATCTTCTCTAAAGTCAAACTGTTGAGTGTATTCAGAGGCCAGTG 960  
DB 901 AATTGCCAGAAATCTTCTCTAAAGTCAAACTGTTGAGTGTATTCAGAGGCCAGTG 960  
QY 961 TAGCGCTTCTAAACCTGAATTAACCCCAAGGATTTCAAGAAATTAAGTAAAGTAAAG 1020  
DB 961 TAGCGCTTCTAAACCTGAATTAACCCCAAGGATTTCAAGAAATTAAGTAAAGTAAAG 1020  
QY 1021 AGACATGACCATGATGGAAGAACTGACATAAGTCTCTGATATTTATTTATGGATC 1080  
DB 1021 AGACATGACCATGATGGAAGAACTGACATAAGTCTCTGATATTTATTTATGGATC 1080  
QY 1081 TGAGCGCCATCAGCGAGTCTTAAGCAGATGAAGACATCTCTGCCAGTCGTTTGAAGATA 1140  
DB 1081 TGAGCGCCATCAGCGAGTCTTAAGCAGATGAAGACATCTCTGCCAGTCGTTTGAAGATA 1140  
QY 1141 ACGAAGAACTGAGGTACTCATTTGCGATCTATCGAGGAGCATGACCATGCGTTTGAAGATA 1200  
DB 1141 ACGAAGAACTGAGGTACTCATTTGCGATCTATCGAGGAGCATGACCATGCGTTTGAAGATA 1200  
QY 1201 TTTTCTATGTCACCAACGGGCGATTTCCATCTCGGCTGACCTTTGACATCTCTCGAGTGA 1260  
DB 1201 TTTTCTATGTCACCAACGGGCGATTTCCATCTCGGCTGACCTTTGACATCTCTCGAGTGA 1260  
QY 1261 CAATAGTAAACACACCGAGGATGTTTTCGAAATTTGAGCCACTTGCCTACCTTTAGTTCAC 1320  
DB 1261 CAATAGTAAACACACCGAGGATGTTTTCGAAATTTGAGCCACTTGCCTACCTTTAGTTCAC 1320  
QY 1321 CTGCTATGGAAGTCACTCATCGCATGGAAGGCTGTCAGAGAGTGTATTTACTATA 1380  
DB 1321 CTGCTATGGAAGTCACTCATCGCATGGAAGGCTGTCAGAGAGTGTATTTACTATA 1380  
QY 1381 ATGATGATGTCATGTTTGGGAAGGATGTTGCGCCAGATGATTTTACAGTCACTCCAAAG 1440  
DB 1381 ATGATGATGTCATGTTTGGGAAGGATGTTGCGCCAGATGATTTTACAGTCACTCCAAAG 1440  
QY 1441 GCGAAGGTTTATTTGACATGCGCTGTCACAACTGTCGCGAGGCTGCCAGGTTCT 1500  
DB 1441 GCGAAGGTTTATTTGACATGCGCTGTCACAACTGTCGCGAGGCTGCCAGGTTCT 1500  
QY 1501 GGAATTAAGGATGCTTATGTCACAGGCTTGTAAATTAATTCAGCTCGGATTTGGATGGTG 1560  
DB 1501 GGAATTAAGGATGCTTATGTCACAGGCTTGTAAATTAATTCAGCTCGGATTTGGATGGTG 1560  
QY 1561 GGAATGCTCTGGAACAGTGGAGGAGTCCGCTATATTCAGAGGTGAGGTTACTGGGA 1620  
DB 1561 GGAATGCTCTGGAACAGTGGAGGAGTCCGCTATATTCAGAGGTGAGGTTACTGGGA 1620  
QY 1621 GTATTTGGAGTGGACACCCCTGCGCTGTTGTTGAGGAAATTAACAGTGTCTCTTACTGTA 1680  
DB 1621 GTATTTGGAGTGGACACCCCTGCGCTGTTGTTGAGGAAATTAACAGTGTCTCTTACTGTA 1680  
QY 1681 ATCAGGAGTGGCAATTCCTGCTCGCTGATAAGTTCGTGACCAAGCATGCAATGTCT 1740  
DB 1681 ATCAGGAGTGGCAATTCCTGCTCGCTGATAAGTTCGTGACCAAGCATGCAATGTCT 1740  
QY 1741 TGTCTGTGGTGTGATGCTGGGACATGTTGGGCAAGATCATTTTCATGATTTGTAAG 1800  
DB 1741 TGTCTGTGGTGTGATGCTGGGACATGTTGGGCAAGATCATTTTCATGATTTGTAAG 1800  
QY 1801 TGATCTCTTCTCCCAACCCAGACTCACTATATTTCCAAAAGGTGAATGCTGCTTATT 1860  
DB 1801 TGATCTCTTCTCCCAACCCAGACTCACTATATTTCCAAAAGGTGAATGCTGCTTATT 1860



QY 1861 TCAGCTTTGAGAGTAGCCAAAAGAGAGTTGAAGTGCCCTAGTAGTGAACAATCCAAATAA 1920  
DB TCAGCTTTGAGAGTAGCCAAAAGAGAGTTGAAGTGCCCTAGTAGTGAACAATCCAAATAA 1920  
QY 1921 TTGACATGCTTTCTATTGCGAACAAGTGGAAACCATCCACCTCATTAATGACAGTGGAA 1980  
DB TTGACATGCTTTCTATTGCGAACAAGTGGAAACCATCCACCTCATTAATGACAGTGGAA 1980  
QY 1981 TGAATGCCACCAACAATACATTTTAATCTCAGCTTTCAAAATACAAACGATGAAGAGTTCA 2040  
DB TGAATGCCACCAACAATACATTTTAATCTCAGCTTTCAAAATACAAACGATGAAGAGTTCA 2040  
QY 2041 AAATGCAGATAACAGTGGAGTGGACACAAAGGAGGACCAAAACCTGAATCTACGGCCC 2100  
DB AAATGCAGATAACAGTGGAGTGGACACAAAGGAGGACCAAAACCTGAATCTACGGCCC 2100  
QY 2101 AGAAGGGTTACGAAAATTTAGTTAGTCCCATAAACATCTTCCAGAGCGGAAATCCTTT 2160  
DB AGAAGGGTTACGAAAATTTAGTTAGTCCCATAAACATCTTCCAGAGCGGAAATCCTTT 2160  
QY 2161 TTGAGGATATTTCCAAAGAAAACCGCTTCCGAAAGTTTAAAGACATGATGTTAACTCAA 2220  
DB TTGAGGATATTTCCAAAGAAAACCGCTTCCGAAAGTTTAAAGACATGATGTTAACTCAA 2220  
QY 2221 CAAGGAGGCGGAGGAGTGAATAATTCCTCTGGTAAATATTTCACTGCTTCCAAAG 2280  
DB CAAGGAGGCGGAGGAGTGAATAATTCCTCTGGTAAATATTTCACTGCTTCCAAAG 2280  
QY 2281 ACGCCAGTTCAGTCTCAATACCTTGGATTTGCAATGCAACATCTTCCAGAGCGGAAATCCTTT 2340  
DB ACGCCAGTTCAGTCTCAATACCTTGGATTTGCAATGCAACATCTTCCAGAGCGGAAATCCTTT 2340  
QY 2341 AAGGATCAATTTGTCGAAGTTCAGCTTGTGAGATCATTTCTGATGAACTCAGCATG 2400  
DB AAGGATCAATTTGTCGAAGTTCAGCTTGTGAGATCATTTCTGATGAACTCAGCATG 2400  
QY 2401 CTAAAAATAAAATCAAGCTATAATAACAGATGAACAAATGACAGTTTGGTCTCCAC 2460  
DB CTAAAAATAAAATCAAGCTATAATAACAGATGAACAAATGACAGTTTGGTCTCCAC 2460  
QY 2461 AGGAAAAACAGGTTTCAAAAAGCATTTGCGAAGCAGCTTAGGAGTGTCTGAAAGATTGC 2520  
DB AGGAAAAACAGGTTTCAAAAAGCATTTGCGAAGCAGCTTAGGAGTGTCTGAAAGATTGC 2520  
QY 2521 AAGGTTGACATTTTCCAGTGAAGTGTAAAGTGAATGGTTCATGACAGGTCAGAAATC 2580  
DB AAGGTTGACATTTTCCAGTGAAGTGTAAAGTGAATGGTTCATGACAGGTCAGAAATC 2580  
QY 2581 CACCCCTGGACTTGGAGACCAAGCAAGATTTAGAGTGGAACTCACACCCAAAAACCA 2640  
DB CACCCCTGGACTTGGAGACCAAGCAAGATTTAGAGTGGAACTCACACCCAAAAACCA 2640  
QY 2641 TAGCGGAAATGTGACAAAAAGAAAGCCCCCATCTCTGATGTTCCACTGGAAGCCAGA 2700  
DB TAGCGGAAATGTGACAAAAAGAAAGCCCCCATCTCTGATGTTCCACTGGAAGCCAGA 2700  
QY 2701 TGACAAAAAGAAAAGAAAATCAAGGAAAGAAAAGAGAACAGTAGAATGGAGGAAATG 2760  
DB TGACAAAAAGAAAAGAAAATCAAGGAAAGAAAAGAGAACAGTAGAATGGAGGAAATG 2760  
QY 2761 CTGAAAATCAATAGGCGGTTACTGAAAGTGTACTTTGGAAGAAAGCTCAGCATTTACAG 2820  
DB CTGAAAATCAATAGGCGGTTACTGAAAGTGTACTTTGGAAGAAAGCTCAGCATTTACAG 2820  
QY 2821 ATAGTTACTTGGGCTTTTTCATGCGGAGAAAAGAGTATTTCCAGATCTTCTGACG 2880  
DB ATAGTTACTTGGGCTTTTTCATGCGGAGAAAAGAGTATTTCCAGATCTTCTGACG 2880  
QY 2881 AAGAGAGTCAATGAAGACAAATTTGGCATTTCACTGATAGCAAAATACTGGAGGC 2940  
DB AAGAGAGTCAATGAAGACAAATTTGGCATTTCACTGATAGCAAAATACTGGAGGC 2940

QY 2941 AAATAAAGATACATTTGCGAGATCCCTCAGATATGTAATAAATAAATTTCTAATAAGCAAGT 3000  
DB AAATAAAGATACATTTGCGAGATCCCTCAGATATGTAATAAATAAATTTCTAATAAGCAAGT 3000  
QY 3001 TTGGAATTCACATCGCGGAAAGTCCCTGCTCATGCTCATGCTCATGATTTGACCGGATTTGTA 3060  
DB TTGGAATTCACATCGCGGAAAGTCCCTGCTCATGCTCATGCTCATGATTTGACCGGATTTGTA 3060  
QY 3061 TGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGCAAGACGCTCATTTCAAAAGTGC 3120  
DB TGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGCAAGACGCTCATTTCAAAAGTGC 3120  
QY 3121 GCATTTCTGAGATATGAGTTGCTTCTCTTATTTTATATCTCATGATGCTCAGTGC 3180  
DB GCATTTCTGAGATATGAGTTGCTTCTCTTATTTTATATCTCATGATGCTCAGTGC 3180  
QY 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGCTGCTTGT 3240  
DB AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGCTGCTTGT 3240  
QY 3241 CTGACAGAGAAATCCGAAACATCTGGCTACAGAAATTCACGAACTGCCGTTAAGTTTGAGG 3300  
DB CTGACAGAGAAATCCGAAACATCTGGCTACAGAAATTCACGAACTGCCGTTAAGTTTGAGG 3300  
QY 3301 ATTTGACAGGCTGGAACACATGCTAATAAATGCTCAAAATGCTTCTGCTGATATCA 3360  
DB ATTTGACAGGCTGGAACACATGCTAATAAATGCTCAAAATGCTTCTGCTGATATCA 3360  
QY 3361 CGCAGCTAAATATATTTCCACCACTCAGGAATCCTACTATGATCCCAACCTGCCACCGG 3420  
DB CGCAGCTAAATATATTTCCACCACTCAGGAATCCTACTATGATCCCAACCTGCCACCGG 3420  
QY 3421 TCACATAAAGTCTAGTAACAACTGTAAACAGTAACCTGACAAATCCACAAAGCATATA 3480  
DB TCACATAAAGTCTAGTAACAACTGTAAACAGTAACCTGACAAATCCACAAAGCATATA 3480  
QY 3481 AGGACAAAAACAAATATAGGTTTGAATCATGGGAGAAAGAAATCGCTTTAAAAATGA 3540  
DB AGGACAAAAACAAATATAGGTTTGAATCATGGGAGAAAGAAATCGCTTTAAAAATGA 3540  
QY 3541 TTCGTACCAAGCTTCTCATGTTGTTGSCCAGTTGGATGACATAGAAAAACCTTAGGA 3600  
DB TTCGTACCAAGCTTCTCATGTTGTTGSCCAGTTGGATGACATAGAAAAACCTTAGGA 3600  
QY 3601 AGTTTGTGCTGCTGAAATGACAAACATTTGACCAATCATATAAGATGCTCAGACAGTGAAG 3660  
DB AGTTTGTGCTGCTGAAATGACAAACATTTGACCAATCATATAAGATGCTCAGACAGTGAAG 3660  
QY 3661 CTGTTCTCAGGACCTTCTATGAAATCCATTTGCCCATACCTTCCCAATTTGAACTGCCAA 3720  
DB CTGTTCTCAGGACCTTCTATGAAATCCATTTGCCCATACCTTCCCAATTTGAACTGCCAA 3720  
QY 3721 GAGAGTATCGAAACCGTTTCTTATGATGCTGAGTGCAGAAATGGAGGCTTATCGAG 3780  
DB GAGAGTATCGAAACCGTTTCTTATGATGCTGAGTGCAGAAATGGAGGCTTATCGAG 3780  
QY 3781 ACAAAATTGAAGTTTGGACCCATTTGTGATAGCAACATTTGATTTATGTTTCTATTTCT 3840  
DB ACAAAATTGAAGTTTGGACCCATTTGTGATAGCAACATTTGATTTATGTTTCTATTTCT 3840  
QY 3841 CATTTTTGCTGAGAGTAAATGCACTTAAGCGGAGATATTTCCAGAGAGGAGTAC 3900  
DB CATTTTTGCTGAGAGTAAATGCACTTAAGCGGAGATATTTCCAGAGAGGAGTAC 3900  
QY 3901 ACAAAAGAGCTAGTCCCAATCGAATCAGAGTATAGAGATCTTCAATTTGAAACCATCTA 3960  
DB ACAAAAGAGCTAGTCCCAATCGAATCAGAGTATAGAGATCTTCAATTTGAAACCATCTA 3960  
QY 3961 CCTCAGCATTTACTGAGCATTTTAAACTCAGCTTCAGAGATGCTTTTGTGATGAT 4020  
DB CCTCAGCATTTACTGAGCATTTTAAACTCAGCTTCAGAGATGCTTTTGTGATGAT 4020  
QY 4021 GCTTAGCAGTTTGGCCCGAAGAAAGAAATATCCAGTACCATGCTGTTTCTTGGCATGAA 4080



XX	(NOVA-) NOVAZYME PHARM INC.	QY	601	CAGCCAAACATCACCTGAAAGACGTCGCATCTCTTTATCTCTTTTCACTTCTGCGAGTG	660
XX	Canfield WM;	DB	601	CAGCCAAACATCACCTGAAAGACGTCGCATCTCTTTATCTCTTTTCACTTCTGCGAGTG	660
XX		QY	661	ACATTTTCAATGTTGCAAAACCAAAACCCCTTACCAATGTCCTCAGTTCTGTTTTTG	720
XX	WPI; 2003-810984/76.	DB	661	ACATTTTCAATGTTGCAAAACCAAAACCCCTTACCAATGTCCTCAGTTCTGTTTTTG	720
XX	P-PSDB; ABW01488, ABW01489.	QY	721	ACAGTACTAAGGATGTTGAAGATGCCACTCTGGACTCTCTGAAATGCTTAAAGGAAATAGCAGACA	780
XX	Producing a high mannose glycoprotein for treating lysosomal storage	DB	721	ACAGTACTAAGGATGTTGAAGATGCCACTCTGGACTCTCTGAAATGCTTAAAGGAAATAGCAGACA	780
XX	PT disease, comprises culturing the lectin resistant mammalian cell in the	QY	781	CAGTATGAGGGGGTACTTGACACAGATTAAGAGTCCCTGGATTAGTGCTTAATGGAAG	840
XX	presence of deoxymannojirimycin and kifunensine.	DB	781	CAGTATGAGGGGGTACTTGACACAGATTAAGAGTCCCTGGATTAGTGCTTAATGGAAG	840
XX	Claim 12; Page 13-15; 46pp; English.	QY	841	ATTTGGCTTTCTGAGTGGATTTCACCAACATTCAGGAAACAAATCACTAAACAA	900
XX	The invention relates to a method for producing a high mannose	DB	841	ATTTGGCTTTCTGAGTGGATTTCACCAACATTCAGGAAACAAATCACTAAACAA	900
XX	glycoprotein. The method comprises: introducing and expressing a	QY	901	AATTCGCAAAATCTTCTCTAAAGTCAAACTGTTCCAGTTGTAATTCAGAGCCAGTG	960
XX	polynucleotide encoding a glycoprotein into a mammalian cell; culturing	DB	901	AATTCGCAAAATCTTCTCTAAAGTCAAACTGTTCCAGTTGTAATTCAGAGCCAGTG	960
XX	the cell in the presence of a lectin to obtain a lectin resistant cell;	QY	961	TAGCGCTTCTAAACTGAATAACCCCAAGATTTTCAAGAAATGAAATAGCAAACTAAGA	1020
XX	isolating the cell; culturing the cell in the presence of	DB	961	TAGCGCTTCTAAACTGAATAACCCCAAGATTTTCAAGAAATGAAATAGCAAACTAAGA	1020
XX	deoxymannojirimycin and kifunensine to inhibit glycosylation of the	QY	1021	AGAAATGACCATGATGGAAGAACTGACATAAGTCTCTGCATATTTATTTGGGATC	1080
XX	glycoprotein; and collecting the glycoprotein. The invention is useful in	DB	1021	AGAAATGACCATGATGGAAGAACTGACATAAGTCTCTGCATATTTATTTGGGATC	1080
XX	gene therapy. The method is useful for producing a high mannose	QY	1081	TGAGCGCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCCAGTCTGTTTGAAGATA	1140
XX	glycoprotein in a complex carbohydrate deficient cell for treating	DB	1081	TGAGCGCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCCAGTCTGTTTGAAGATA	1140
XX	lysosomal storage disease. The present sequence is human N-	QY	1141	ACGAAGAACTGAGTACTCATTTGGATCTATCGAGAGCATGACCATGGTTCGGAATA	1200
XX	acetylglucosamine-1 (GlcNAc)-phosphotransferase alpha/beta subunit	DB	1141	ACGAAGAACTGAGTACTCATTTGGATCTATCGAGAGCATGACCATGGTTCGGAATA	1200
XX	precursor CDNA	QY	1201	TTTTCAATGTCACCAACCGGCGAGTTCCATCTGGCTGAACTTGACATCTCTCGAGTGA	1260
XX	Sequence 5597 BP; 1678 A; 1151 C; 1219 G; 1549 T; 0 U; 0 Other;	DB	1201	TTTTCAATGTCACCAACCGGCGAGTTCCATCTGGCTGAACTTGACATCTCTCGAGTGA	1260
XX	Query Match 100.0%; Score 5597; DB 10; Length 5597;	QY	1261	CAATAGTAAACACACACAGGATGTTTTTCGAAATTTGAGCCACTTGCCTTACCTTACCT	1320
XX	Best Local Similarity 100.0%; Pred. No. 0;	DB	1261	CAATAGTAAACACACACAGGATGTTTTTCGAAATTTGAGCCACTTGCCTTACCTTACCT	1320
XX	Matches 5597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	QY	1321	CTGCTATTGAAAGTCACATTCATCGCATCGAAGGGCTGTCGCAAGTTTATTTACCTAA	1380
QY	1 CGGAGCCGAGCGGGGTCCTGTCGGAGCTGCAATGAGGGGCGCCGCGAGGCTGTGACC	DB	1321	CTGCTATTGAAAGTCACATTCATCGCATCGAAGGGCTGTCGCAAGTTTATTTACCTAA	1380
DB	1 CGGAGCCGAGCGGGGTCCTGTCGGAGCTGCAATGAGGGGCGCCGCGAGGCTGTGACC	QY	1381	ATGATGATGTCATGTTTGGGAAAGATGTCGCGCAGATGATTTTACAGTCACCTCCAAAG	1440
QY	61 TGGCGCGCGCGCGCGCGCGCGCTGTAATGCGCGCTGCTGAGCGCGCGCGCGCGC	DB	1381	ATGATGATGTCATGTTTGGGAAAGATGTCGCGCAGATGATTTTACAGTCACCTCCAAAG	1440
DB	61 TGGCGCGCGCGCGCGCGCGCGCTGTAATGCGCGCTGCTGAGCGCGCGCGCGCGC	QY	1441	GCCAGAAAGTTATTTGACATGGCTGTGCCAACTGTGCCGAGGGCTGCCAGGTTCTCT	1500
QY	121 GCGGCTCAGGCTCTCTGCGGCGTGGCGTGGCGTGAAGGGTGAATCTTTCAAGTCTC	DB	1441	GCCAGAAAGTTATTTGACATGGCTGTGCCAACTGTGCCGAGGGCTGCCAGGTTCTCT	1500
DB	121 GCGGCTCAGGCTCTCTGCGGCGTGGCGTGGCGTGAAGGGTGAATCTTTCAAGTCTC	QY	1501	GGATTAAGGATGCGCTTATTTGACAGGCTTCTTAATTTACCTGCGATTTGGATGGTG	1560
QY	181 TGCAGAGAACACTATACCTGCTGTCGCCACAGGTATGGCTCTACGTCTGCTTCTTG	DB	1501	GGATTAAGGATGCGCTTATTTGACAGGCTTCTTAATTTACCTGCGATTTGGATGGTG	1560
DB	181 TGCAGAGAACACTATACCTGCTGTCGCCACAGGTATGGCTCTACGTCTGCTTCTTG	QY	1561	GGGATTGCTCTGGAACAGTGGAGGAGTCTATTTATTCAGGAGGTGGAGGTACTGGGA	1620
QY	241 GCGTCTGTTGTCACCATGCTCTCCGCTTCCAGTTCGGAGAGTGGTTCTGGAATGGAGCC	DB	1561	GGGATTGCTCTGGAACAGTGGAGGAGTCTATTTATTCAGGAGGTGGAGGTACTGGGA	1620
DB	241 GCGTCTGTTGTCACCATGCTCTCCGCTTCCAGTTCGGAGAGTGGTTCTGGAATGGAGCC	QY	1621	GTATTTGGATTTGGACACCCCTGGCAGTTTGGTGGAGGATTAACAGTGTCTCTTACTGTA	1680
QY	301 GAGTCAATACCATGTTTGTGTTGATTCCTATAGACAATATTTGCTGGAAGTCTTTTC	DB	1621	GTATTTGGATTTGGACACCCCTGGCAGTTTGGTGGAGGATTAACAGTGTCTCTTACTGTA	1680
DB	301 GAGTCAATACCATGTTTGTGTTGATTCCTATAGACAATATTTGCTGGAAGTCTTTTC	QY	1681	ATCAGGGATGTCGCAATGCTTAAAGTGGCAATGCTTGTACTGACCCAGCCCTGC	1740
QY	361 AGAATCGGCTTTGCTGCCCATGCGGATGACGTTGTTTACACCTGGGTGAATGGCACAG	DB			
DB	361 AGAATCGGCTTTGCTGCCCATGCGGATGACGTTGTTTACACCTGGGTGAATGGCACAG	QY			
QY	421 ATCTTGAACACTAGAGGAACCTACAGAGGTTCAGAAACAGATGGAGGAGGACAGAAAG	DB			
DB	421 ATCTTGAACACTAGAGGAACCTACAGAGGTTCAGAAACAGATGGAGGAGGACAGAAAG	QY			
QY	481 CAATGAGAGAATCTCTTGGGAAAACCAACCGAACCTACTAAGAGAGTGAGAAGCAGT	DB			
DB	481 CAATGAGAGAATCTCTTGGGAAAACCAACCGAACCTACTAAGAGAGTGAGAAGCAGT	QY			
QY	541 TAGAGTGTTCCTAACACACTGCAATTAAGGTGCGCAATGCTGTACTGGAACCCAGCCCTGC	DB			
DB	541 TAGAGTGTTCCTAACACACTGCAATTAAGGTGCGCAATGCTGTACTGGAACCCAGCCCTGC				

Db 1681 ATCAGGGATGTCGAAATTCCTGGCTCGCTGATAAGTTCGTGACCAAGCATGCAATGCT 1740  
Qy 1741 TGTCTGTGGTTCATCTGCTGGCGACTGTGGCGAAGATCATTTTCATGAATGTATTAAG 1800  
Db 1741 TGTCTGTGGTTCATCTGCTGGCGACTGTGGCGAAGATCATTTTCATGAATGTATTAAG 1800  
Qy 1801 TGATCCTTCTCCAAACAGAGCTCACTATATATTATCCAAAAGGTGAATGCCCTGCTTATT 1860  
Db 1801 TGATCCTTCTCCAAACAGAGCTCACTATATATTATCCAAAAGGTGAATGCCCTGCTTATT 1860  
Qy 1861 TCAGCTTTGCGAAGTAGCCAAAGAGAGGTGAAGTGGCTATAGTGACAATCAATCAATAA 1920  
Db 1861 TCAGCTTTGCGAAGTAGCCAAAGAGAGGTGAAGTGGCTATAGTGACAATCAATCAATAA 1920  
Qy 1921 TTCGACATGCTTCTATTCGCAACAGAGTGGAACCATCCACTCATATATGACAGTGGAA 1980  
Db 1921 TTCGACATGCTTCTATTCGCAACAGAGTGGAACCATCCACTCATATATGACAGTGGAA 1980  
Qy 1981 TGAATGCCACCAACATATTTTAACTCAGCTTTCAAAATACAAACGATGAAGAGTTCA 2040  
Db 1981 TGAATGCCACCAACATATTTTAACTCAGCTTTCAAAATACAAACGATGAAGAGTTCA 2040  
Qy 2041 AAATGCAGATACAGTGAGGTGGACACAAGGGAGGGACCAAACTGAAATTTACGGCCC 2100  
Db 2041 AAATGCAGATACAGTGAGGTGGACACAAGGGAGGGACCAAACTGAAATTTACGGCCC 2100  
Qy 2101 AGAAGGGTTAGCAAAATTTAGTTAGTCCCATAAACACTTCTCCAGAGCGGAAATCCCTT 2160  
Db 2101 AGAAGGGTTAGCAAAATTTAGTTAGTCCCATAAACACTTCTCCAGAGCGGAAATCCCTT 2160  
Qy 2161 TTGAGGATATTCGCAAGAAAACGCTTCCGAGTTTAAGAGACATGATGTAACTCAA 2220  
Db 2161 TTGAGGATATTCGCAAGAAAACGCTTCCGAGTTTAAGAGACATGATGTAACTCAA 2220  
Qy 2221 CAAGGAGAGCCAGGAGAGGTGAAATTCCTCGTGAATATTTCACTCTCTCCAAAG 2280  
Db 2221 CAAGGAGAGCCAGGAGAGGTGAAATTCCTCGTGAATATTTCACTCTCTCCAAAG 2280  
Qy 2281 ACGCCAGTGTAGTCTCAATACCTTGGATTTGCAACTGGAAACATGAGACATCACTTTGA 2340  
Db 2281 ACGCCAGTGTAGTCTCAATACCTTGGATTTGCAACTGGAAACATGAGACATCACTTTGA 2340  
Qy 2341 AAGGATCAATTTGTCGAAGTCAGCTTGTGAGATCATTTCTGATGAACTCAGCAGTG 2400  
Db 2341 AAGGATCAATTTGTCGAAGTCAGCTTGTGAGATCATTTCTGATGAACTCAGCAGTG 2400  
Qy 2401 CTAAATATAAAATCAAGCTATATAACAGATGAAACAAATGACAGTTTGGTGGCTCCAC 2460  
Db 2401 CTAAATATAAAATCAAGCTATATAACAGATGAAACAAATGACAGTTTGGTGGCTCCAC 2460  
Qy 2461 AGGAAAAACAGTTTCATATAAGCATCTTGCCAAACAGCTTAGGAGTGTCTGAAAAGATTGC 2520  
Db 2461 AGGAAAAACAGTTTCATATAAGCATCTTGCCAAACAGCTTAGGAGTGTCTGAAAAGATTGC 2520  
Qy 2521 AGAGTTTCACTTTTCTCCAGTGTGTAAGTGTGATGTCATGACAGGGTCCAGATC 2580  
Db 2521 AGAGTTTCACTTTTCTCCAGTGTGTAAGTGTGATGTCATGACAGGGTCCAGATC 2580  
Qy 2581 CACCCCTCGACTTGAGACCAAGCAAGATTTAGAGTGGAACTCACACCCAAAAACCA 2640  
Db 2581 CACCCCTCGACTTGAGACCAAGCAAGATTTAGAGTGGAACTCACACCCAAAAACCA 2640  
Qy 2641 TAGGGGAATGTGCAAAAGAAAGCCCATCTCTGATTTTCCACTGGAAAGCCAGA 2700  
Db 2641 TAGGGGAATGTGCAAAAGAAAGCCCATCTCTGATTTTCCACTGGAAAGCCAGA 2700  
Qy 2701 TGACAAAAAGAAAGAAATCAAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2760  
Db 2701 TGACAAAAAGAAAGAAATCAAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2760  
Qy 2761 CTGAAATCAGATAGGCTTACTGAGTGTACTTGGNAGAAAGCTGCAGCATTCACAG 2820

Db 2761 CTGAAATCAGATAGGCTTACTGAAGTGTACTTGGAAAGAAAGCTGCAGCATTCACAG 2820  
Qy 2821 ATAGTTTACTTGGCTTTTGGCCATGGGAGAAAGAAAGATTTCCAGAGATCTTCTCGAG 2880  
Db 2821 ATAGTTTACTTGGCTTTTGGCCATGGGAGAAAGAAAGATTTCCAGAGATCTTCTCGAG 2880  
Qy 2881 AAGAAGAGTCAATTGAAGACACAATTTGGCATACTTCACTGATAGCAAAATATCTGGAGGC 2940  
Db 2881 AAGAAGAGTCAATTGAAGACACAATTTGGCATACTTCACTGATAGCAAAATATCTGGAGGC 2940  
Qy 2941 AACTAAAGATACATTTGAGAGTCCCTCAGATATGTAAATTAATTTCTAAATAGCAAGT 3000  
Db 2941 AACTAAAGATACATTTGAGAGTCCCTCAGATATGTAAATTAATTTCTAAATAGCAAGT 3000  
Qy 3001 TTGATTTCACTCGCGGAAAGTCCCTGCTCACATGCTCACATGATTTGACCGGATTTGA 3060  
Db 3001 TTGATTTCACTCGCGGAAAGTCCCTGCTCACATGCTCACATGATTTGACCGGATTTGA 3060  
Qy 3061 TGAAGAACTGCAAGATATGTTCCCTGGAAGAAATTTGACAAGAGCTCATTTCAAAAAGTGC 3120  
Db 3061 TGAAGAACTGCAAGATATGTTCCCTGGAAGAAATTTGACAAGAGCTCATTTCAAAAAGTGC 3120  
Qy 3121 GCCATTTGAGGATATGCAAGTTCCTTCTCTATTTTATTTATCTCATGAGTGCAGTGC 3180  
Db 3121 GCCATTTGAGGATATGCAAGTTCCTTCTCTATTTTATTTATCTCATGAGTGCAGTGC 3180  
Qy 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGCTGTCTGT 3240  
Db 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGCTGTCTGT 3240  
Qy 3241 CTGACAGAGAAATCCGAACTGCTTACCAGATTTCCAGAACTGCGGTTAAGTTTCAGG 3300  
Db 3241 CTGACAGAGAAATCCGAACTGCTTACCAGATTTCCAGAACTGCGGTTAAGTTTCAGG 3300  
Qy 3301 ATTTGACAGTCTGGAACACATGCTAAATAATTTGCTCAAAAATGCTTCTGCTGATATCA 3360  
Db 3301 ATTTGACAGTCTGGAACACATGCTAAATAATTTGCTCAAAAATGCTTCTGCTGATATCA 3360  
Qy 3361 CGCAGCTAAATATATTTCCACAACTCAGGAAATCTTATGATGCCCACTGCGCACCGG 3420  
Db 3361 CGCAGCTAAATATATTTCCACAACTCAGGAAATCTTATGATGCCCACTGCGCACCGG 3420  
Qy 3421 TCCTAAAGTCTAGTAAACAACTGTAACCTGTAACCTGTAACCTGTAACCTGTAACCTGTA 3480  
Db 3421 TCCTAAAGTCTAGTAAACAACTGTAACCTGTAACCTGTAACCTGTAACCTGTAACCTGTA 3480  
Qy 3481 AGGACAAAACAAATATAGGTTTGAATCATGGGAGAAAGAAATTCGTTTTAAATGA 3540  
Db 3481 AGGACAAAACAAATATAGGTTTGAATCATGGGAGAAAGAAATTCGTTTTAAATGA 3540  
Qy 3541 TTGTTACCAAGCTTCTCATGTTGGTGGCCAGTTGGATGACATAAGAAAACCCCTAGGA 3600  
Db 3541 TTGTTACCAAGCTTCTCATGTTGGTGGCCAGTTGGATGACATAAGAAAACCCCTAGGA 3600  
Qy 3601 AGTTTGTGCTGAAATGACAACTGACCAATCATAAAGATGCTCAGACAGTGAAG 3660  
Db 3601 AGTTTGTGCTGAAATGACAACTGACCAATCATAAAGATGCTCAGACAGTGAAG 3660  
Qy 3661 CTGTTCTAGGAGTCTTATGAAATCCATGTTCCCATACCTCCCAATTTGAACCTGCAA 3720  
Db 3661 CTGTTCTAGGAGTCTTATGAAATCCATGTTCCCATACCTCCCAATTTGAACCTGCAA 3720  
Qy 3721 GAGAGTATCGAAACCGTTTCTCATATGATGAGCTGAGGAAATGGAGGGCTTATCGAG 3780  
Db 3721 GAGAGTATCGAAACCGTTTCTCATATGATGAGCTGAGGAAATGGAGGGCTTATCGAG 3780  
Qy 3781 ACAAATGAAGTTTGGAGCCCATTTGTTACTAGCAACATTTGATTTATTTTACTATTTCT 3840  
Db 3781 ACAAATGAAGTTTGGAGCCCATTTGTTACTAGCAACATTTGATTTATTTTACTATTTCT 3840  
Qy 3841 CATTTTTTCTGAGCAGTTTAAATTTGCACTTAAAGCGGAAAGATTTTCCAGAGAGGATAC 3900  
Db 3841 CATTTTTTCTGAGCAGTTTAAATTTGCACTTAAAGCGGAAAGATTTTCCAGAGAGGATAC 3900

QY	3901	ACAAAGAGCTAGTCCCAATCGAATCAGAGTAGAAGATCTTTCATTTGAAAAACCATCTA	3960
DB	3901	ACAAAGAGCTAGTCCCAATCGAATCAGAGTAGAAGATCTTTCATTTGAAAAACCATCTA	3960
QY	3961	CTTCAGCAATTTCTGAGCATTTTAAACTCAGCTTCACAGAGATGCTTTGTGATGTGAT	4020
DB	3961	CTTCAGCAATTTCTGAGCATTTTAAACTCAGCTTCACAGAGATGCTTTGTGATGTGAT	4020
QY	4021	GCTTAGCAGTTTGGCCCGAAGAGGAAAATATCCAGTACCATCGTGTGTTTTGTGCGCATGAA	4080
DB	4021	GCTTAGCAGTTTGGCCCGAAGAGGAAAATATCCAGTACCATCGTGTGTTTTGTGCGCATGAA	4080
QY	4081	TATAGCCCACTGACTAGGAATTTTAAACCAACCCACTGTAATACTTGCTGTCGAGCAGC	4140
DB	4081	TATAGCCCACTGACTAGGAATTTTAAACCAACCCACTGTAATACTTGCTGTCGAGCAGC	4140
QY	4141	TCTGAACGTGATTTTACTTTTAAAGAAATTTGCTCATCGAGCTGTCACTCTTTTATATAAAA	4200
DB	4141	TCTGAACGTGATTTTACTTTTAAAGAAATTTGCTCATCGAGCTGTCACTCTTTTATATAAAA	4200
QY	4201	GGCTCACTGCAAGAGACAGCTGTTAAATTTCCACAGCAATCATTTGCAGACTAACTTTAT	4260
DB	4201	GGCTCACTGCAAGAGACAGCTGTTAAATTTCCACAGCAATCATTTGCAGACTAACTTTAT	4260
QY	4261	TAGAGAAAGCCTATGCCAGCTGGAGTGATGCTTAGAGGCTCCAGTCTTTGCAATCCAA	4320
DB	4261	TAGAGAAAGCCTATGCCAGCTGGAGTGATGCTTAGAGGCTCCAGTCTTTGCAATCCAA	4320
QY	4321	AGCCTTTTGCTAAAGATTTTGCACCTTTTTTTTTTTTCATTTCCCATTTTAAAGTAGTTACTA	4380
DB	4321	AGCCTTTTGCTAAAGATTTTGCACCTTTTTTTTTTTTCATTTCCCATTTTAAAGTAGTTACTA	4380
QY	4381	AGTTAACTAGTATTTCTTGCTTCGTAGTATAACGAATGGGATGCTTAAACCTTATTTTAA	4440
DB	4381	AGTTAACTAGTATTTCTTGCTTCGTAGTATAACGAATGGGATGCTTAAACCTTATTTTAA	4440
QY	4441	TAGATGTTATTTTAAATAATCAGCAATATCACCTCTTATTGCAATACCTTAAATTTATGAG	4500
DB	4441	TAGATGTTATTTTAAATAATCAGCAATATCACCTCTTATTGCAATACCTTAAATTTATGAG	4500
QY	4501	TTTTATTAAATTTAAGACTGTAAATGGCTTTAAACCACTAATCTGTAAGAGCTCAATG	4560
DB	4501	TTTTATTAAATTTAAGACTGTAAATGGCTTTAAACCACTAATCTGTAAGAGCTCAATG	4560
QY	4561	ATTGACATCTGAATGCTTTTGTAAATTTAGACTTTAGCCCCCTAAGAACTGATGATTTCA	4620
DB	4561	ATTGACATCTGAATGCTTTTGTAAATTTAGACTTTAGCCCCCTAAGAACTGATGATTTCA	4620
QY	4621	CGTGCAGGCTCTAATTTTCAACAGGCTAGAGTAGTAGTACTTACCAGATGTAATATGTTT	4680
DB	4621	CGTGCAGGCTCTAATTTTCAACAGGCTAGAGTAGTAGTACTTACCAGATGTAATATGTTT	4680
QY	4681	TGGAAATGTACATATTTCAAAACAGAGTGCCTCATTTTAGAAATGAGTAGTGTATGGCA	4740
DB	4681	TGGAAATGTACATATTTCAAAACAGAGTGCCTCATTTTAGAAATGAGTAGTGTATGGCA	4740
QY	4741	CTGSCAATTTACAGTGGTGTCTTTGTTTAAATCTCATTGGTATATTCAGTAGCTATCTCT	4800
DB	4741	CTGSCAATTTACAGTGGTGTCTTTGTTTAAATCTCATTGGTATATTCAGTAGCTATCTCT	4800
QY	4801	CTCAGTTGGTTTTTGTATAGAACAGAGCCAGCAAACTTTCTTTGTAAAAGCGTGGTTAGT	4860
DB	4801	CTCAGTTGGTTTTTGTATAGAACAGAGCCAGCAAACTTTCTTTGTAAAAGCGTGGTTAGT	4860
QY	4861	AAATATTGCAGGCCACCTGTGCTTTTGTGCATACATTTCTTTGCTGTGTTTGTAGTTGT	4920
DB	4861	AAATATTGCAGGCCACCTGTGCTTTTGTGCATACATTTCTTTGCTGTGTTTGTAGTTGT	4920
QY	4921	TTTTTTTTCAAAACCAACCTCTTAAATAATGTAAAAACCAATGTTTAGCTTGCAGCTGTACAAA	4980
DB	4921	TTTTTTTTCAAAACCAACCTCTTAAATAATGTAAAAACCAATGTTTAGCTTGCAGCTGTACAAA	4980

QY	4981		ACTGCCACCAGCCAGATGTGACCTTCAGSCCACTCATTTGGCCAATCAGTGAGAATTATTTT	5040
Db	4981		ACTGCCCAACAGCCAGATGTGACCCCTCAGGCCATCATTTGGCCAATCACGTAGAATTATTTT	5040
QY	5041		T TGTTTGTTGTTTGTTTGTTTGTTTGAGACAGAGTCTCTCTCTGTTGCCCGAGCTGGAG	5100
Db	5041		T TGTTTGTTGTTTGTTTGTTTGTTTGAGACAGAGTCTCTCTCTGTTGCCCGAGCTGGAG	5100
QY	5101		TGCAGTGGCGCAATCTCAGCTCAGTCCAACTCCGGCTCCC GGSTTCAAAGCAGTCTCTGC	5160
Db	5101		TGCAGTGGCGCAATCTCAGCTCAGTCCAACTCCGGCTCCC GGSTTCAAAGCAGTCTCTGC	5160
QY	5161		TCAGCCTTCAGTAGTGGGATACAGGTGCATGCCACACACCCCTGCTAATTTTTTGA	5220
Db	5161		TCAGCCTTCAGTAGTGGGATACAGGTGCATGCCACACACCCCTGCTAATTTTTTGA	5220
QY	5221		TTTTTAGTAGAGCGGGGTTCCACCATATTTGCTCAGGCTTATCTTGAACCTCCTGACCTC	5280
Db	5221		TTTTTAGTAGAGCGGGGTTCCACCATATTTGCTCAGGCTTATCTTGAACCTCCTGACCTC	5280
QY	5281		AGGTGATCCACCTGCCCTTCGCCCTCCAAAAGTGTGAGATTAAGGCATTAAGCCAGTGCAC	5340
Db	5281		AGGTGATCCACCTGCCCTTCGCCCTCCAAAAGTGTGAGATTAAGGCATTAAGCCAGTGCAC	5340
QY	5341		CCAGCCGAGAAATAGTATTTTTTATGTATGTTAAACCTTGGCGTCTAGCCATAATTTATG	5400
Db	5341		CCAGCCGAGAAATAGTATTTTTTATGTATGTTAAACCTTGGCGTCTAGCCATAATTTATG	5400
QY	5401		TCATAATACAATGGATTTGTGAAGACAGATTC CATGAGTAACTCTGACAGTATTTTATG	5460
Db	5401		TCATAATACAATGGATTTGTGAAGACAGATTC CATGAGTAACTCTGACAGTATTTTATG	5460
QY	5461		ATCATGATCTCAACATATCTCTCCCAATGGCATACATCTTTTGTAACAAGAACCTTGAA	5520
Db	5461		ATCATGATCTCAACATATCTCTCCCAATGGCATACATCTTTTGTAACAAGAACCTTGAA	5520
QY	5521		ATGTAATACTGTGTTTGCTGTGAAGAGTGTGTAATTTCAAAAACTGAAATCTCATAAA	5580
Db	5521		ATGTAATACTGTGTTTGCTGTGAAGAGTGTGTAATTTCAAAAACTGAAATCTCATAAA	5580
QY	5581		AAGTTAAATTTTGA AAA 5597	
Db	5581		AAGTTAAATTTTGA AAA 5597	
RESULT 4				
ID	AAD62650	standard; cDNA; 5597 BP.		
XX AC	AAD62650;			
DT DT	15-JAN-2004	(first entry)		
XX DE	Human G1cNAc-phosphotransferase alpha/beta subunit precursor cdna.			
XX DX	Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;			
KW KW	N-acetylglucosamine-1-phosphotransferase; gene; gene therapy; human; ss.			
OS OS	Homo sapiens.			
PH FH	Key Location/Qualifiers			
CDS CDS	165..2948			
FT FT	/tag= a			
FT FT	/product= "N-acetylglucosamine-1(GlcNAc) -			
FT FT	phosphotransferase alpha subunit"			
FT FT	/note= "No stop codon"			
FT FT	/partial			
FT FT	2949..3935			
FT FT	/tag= b			
FT FT	/product= "N-acetylglucosamine-1(GlcNAc) -			
FT FT	phosphotransferase beta subunit"			
FT FT	/note= "No start codon"			
FT FT	/partial			





Db 1621 GTATTGGAGTTGGACACCCCTTGGCGAGTTTGGTGGAGGNAATAAACAGTGTCTCTTACTGTGA 1680  
Qy 1681 ATCAGGGATGTCGGAATTTCTGGCTCGCTGATAGTTTCTGTGACCAAGCATGCAATCTCT 1740  
Db 1681 ATCAGGGATGTCGGAATTTCTGGCTCGCTGATAGTTTCTGTGACCAAGCATGCAATCTCT 1740  
Qy 1741 TGTCTGTGGTGTGATCTGGCGACCTGGCGCAGATCAATTTTCATGAAATGATTAAG 1800  
Db 1741 TGTCTGTGGTGTGATCTGGCGACCTGGCGCAGATCAATTTTCATGAAATGATTAAG 1800  
Qy 1801 TGATCTTCTCCCAACAGACTCACTATATTTATTTCCAAAGAGTGAATGCTGCTTATT 1860  
Db 1801 TGATCTTCTCCCAACAGACTCACTATATTTATTTCCAAAGAGTGAATGCTGCTTATT 1860  
Qy 1861 TCAGCTTTGCGAAGTAGCCAAAGAGGAGTTGAAGTGGCTATAGTGACAATCCAATAA 1920  
Db 1861 TCAGCTTTGCGAAGTAGCCAAAGAGGAGTTGAAGTGGCTATAGTGACAATCCAATAA 1920  
Qy 1921 TTCCGACATGCTTCTATTTGCCAACAGTGGAAAAACATCCACCTCATATGCAAGTGGAA 1980  
Db 1921 TTCCGACATGCTTCTATTTGCCAACAGTGGAAAAACATCCACCTCATATGCAAGTGGAA 1980  
Qy 1981 TGAATGCCACCAATACATTTTAACTTCACTGTTTCAAAATPACAAACGATGAAGATPCA 2040  
Db 1981 TGAATGCCACCAATACATTTTAACTTCACTGTTTCAAAATPACAAACGATGAAGATPCA 2040  
Qy 2041 AAATGSCAGATAACAGTGGAGTGGACACCAAGGAGGACCAAACTCAATTTCTACGGCCC 2100  
Db 2041 AAATGSCAGATAACAGTGGAGTGGACACCAAGGAGGACCAAACTCAATTTCTACGGCCC 2100  
Qy 2101 AGAAGGTTACGAAAAATTTAGTTAGTCCCAATAACACTTTTCCAGAGGCGGAAATCCTTT 2160  
Db 2101 AGAAGGTTACGAAAAATTTAGTTAGTCCCAATAACACTTTTCCAGAGGCGGAAATCCTTT 2160  
Qy 2161 TTGAGGATATTTCCCAAGAAAAAGCTTCCCGAAGTTTAAAGACATGATTTAACTCAA 2220  
Db 2161 TTGAGGATATTTCCCAAGAAAAAGCTTCCCGAAGTTTAAAGACATGATTTAACTCAA 2220  
Qy 2221 CAAGGAGAGCCAGGAAGAGTGAATAATTCCTCCCTGGTAAATATTTTCACTCCTTCCAAAAG 2280  
Db 2221 CAAGGAGAGCCAGGAAGAGTGAATAATTCCTCCCTGGTAAATATTTTCACTCCTTCCAAAAG 2280  
Qy 2281 ACGCCAGTGTAGTCTCAATACCTTGGATTTGCAACTGGAACTGGAGACATCACTTTGA 2340  
Db 2281 ACGCCAGTGTAGTCTCAATACCTTGGATTTGCAACTGGAACTGGAGACATCACTTTGA 2340  
Qy 2341 AAGGATACAAATTTGTCGAAGTCAGCTTGTGAGATCAATTTCTGATGAATCTCAGCATG 2400  
Db 2341 AAGGATACAAATTTGTCGAAGTCAGCTTGTGAGATCAATTTCTGATGAATCTCAGCATG 2400  
Qy 2401 CTAAAAATAAAATCAAGCTATATAACAGATGAAAACAAATGACAGTTTGGTGGCTCCAC 2460  
Db 2401 CTAAAAATAAAATCAAGCTATATAACAGATGAAAACAAATGACAGTTTGGTGGCTCCAC 2460  
Qy 2461 AGGAAAAACAGTTTCAATAAAGACATTTGCCAAGCTTGGAGTGTCTGAAAGATTGC 2520  
Db 2461 AGGAAAAACAGTTTCAATAAAGACATTTGCCAAGCTTGGAGTGTCTGAAAGATTGC 2520  
Qy 2521 AGAGTTTGAATTTTCTCGAGTGTGTAAGTGAATGTTGATGACAGGAGTGTGTAAGTGC 2580  
Db 2521 AGAGTTTGAATTTTCTCGAGTGTGTAAGTGAATGTTGATGACAGGAGTGTGTAAGTGC 2580  
Qy 2581 CACCCCTGGAATTTGACAAAAAGAAAGCCCTCATCTCTGATTTGTTTCCACTGGAAGCCAGA 2640  
Db 2581 CACCCCTGGAATTTGACAAAAAGAAAGCCCTCATCTCTGATTTGTTTCCACTGGAAGCCAGA 2640  
Qy 2641 TAGCGGAAATTTGACAAAAAGAAAGCCCTCATCTCTGATTTGTTTCCACTGGAAGCCAGA 2700  
Db 2641 TAGCGGAAATTTGACAAAAAGAAAGCCCTCATCTCTGATTTGTTTCCACTGGAAGCCAGA 2700  
Qy 2701 TGACAAAAAGAAAGAAATTCACAGGGAAGAAAGAAAGAGACAGTAGAATGAGGAAAGATG 2760  
Db 2701 TGACAAAAAGAAAGAAATTCACAGGGAAGAAAGAAAGAGACAGTAGAATGAGGAAAGATG 2760

Qy 2761 CTGAAAAATCACATAGCGGTTACTGAAGTGTACTTTGGAGAAAGCTGCGAGCATTTACACAG 2820  
Db 2761 CTGAAAAATCACATAGCGGTTACTGAAGTGTACTTTGGAGAAAGCTGCGAGCATTTACACAG 2820  
Qy 2821 ATAGTTTACTTTGGCTTTTGGCCATGGGAGAAAAAGTATTTTCCAGATCTTCTCGACG 2880  
Db 2821 ATAGTTTACTTTGGCTTTTGGCCATGGGAGAAAAAGTATTTTCCAGATCTTCTCGACG 2880  
Qy 2881 AAGAGAGTCATTGGAAGACACAATTTGGCATACTTCACTGATAGCAAAAAATCTGGAGGC 2940  
Db 2881 AAGAGAGTCATTGGAAGACACAATTTGGCATACTTCACTGATAGCAAAAAATCTGGAGGC 2940  
Qy 2941 AACTAAAAAGATACATTTGCGAGATTTCCCTCAGATATGTAAATAAAATTTCTAAATAGCAAGT 3000  
Db 2941 AACTAAAAAGATACATTTGCGAGATTTCCCTCAGATATGTAAATAAAATTTCTAAATAGCAAGT 3000  
Qy 3001 TTGGATTTCACTCGCGGAAAGTCCCTGCTCACTGCTCCTCACTGATTTGACCGGATTGTTA 3060  
Db 3001 TTGGATTTCACTCGCGGAAAGTCCCTGCTCACTGCTCCTCACTGATTTGACCGGATTGTTA 3060  
Qy 3061 TGCAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAGACGTCATTTCAAAAGTGC 3120  
Db 3061 TGCAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAGACGTCATTTCAAAAGTGC 3120  
Qy 3121 GCCATTCTGAGGATATGAGTTTGGCTTCTCTTATTTTATTTTATCTCATGATGTCAGTGC 3180  
Db 3121 GCCATTCTGAGGATATGAGTTTGGCTTCTCTTATTTTATTTTATCTCATGATGTCAGTGC 3180  
Qy 3181 AGCCACTGAATATATCTCAAGTCTTTTGATGAAGTTGATACAGATCAATCTGGTCTTGT 3240  
Db 3181 AGCCACTGAATATATCTCAAGTCTTTTGATGAAGTTGATACAGATCAATCTGGTCTTGT 3240  
Qy 3241 CTGACAGAGAAATCGAACAACCTGGCTACAGAAATTCAGAACTGCCGTTAAGTTTGAGG 3300  
Db 3241 CTGACAGAGAAATCGAACAACCTGGCTACAGAAATTCAGAACTGCCGTTAAGTTTGAGG 3300  
Qy 3301 ATTTGACAGGTTGGAACACATGCTTAATAAATGCTCAAAAATGCTTCTCTGCTGATATCA 3360  
Db 3301 ATTTGACAGGTTGGAACACATGCTTAATAAATGCTCAAAAATGCTTCTCTGCTGATATCA 3360  
Qy 3361 CGCAGCTAAAATAATTTCCACCACCTCAGGAATCTTACTATGATCCCAACCTGCCACCGG 3420  
Db 3361 CGCAGCTAAAATAATTTCCACCACCTCAGGAATCTTACTATGATCCCAACCTGCCACCGG 3420  
Qy 3421 TCACCTAAAAAGTCTAGTAAACAACTGTAAACCCAGTAACTGACAAAAATCCCAAGCATATA 3480  
Db 3421 TCACCTAAAAAGTCTAGTAAACAACTGTAAACCCAGTAACTGACAAAAATCCCAAGCATATA 3480  
Qy 3481 AGGACAAAAACAAATATAGTTTGAATCATGGGAGAGAGAAATCGCTTTTAAATGA 3540  
Db 3481 AGGACAAAAACAAATATAGTTTGAATCATGGGAGAGAGAGAAATCGCTTTTAAATGA 3540  
Qy 3541 TTCGTTACCAACGTTTCTCATGTTGGCCAGTTGGATGACATAGAAAAAACCCCTAGGA 3600  
Db 3541 TTCGTTACCAACGTTTCTCATGTTGGCCAGTTGGATGACATAGAAAAAACCCCTAGGA 3600  
Qy 3601 AGTTTGTGCTGATGACAAACATTTGACCAATCATATAAGATGCTCAGACAGTGAAGG 3660  
Db 3601 AGTTTGTGCTGATGACAAACATTTGACCAATCATATAAGATGCTCAGACAGTGAAGG 3660  
Qy 3661 CTGTTCTCAGGAGCTTCTATGAATCCCATGTTCCCATACCTTCCCAATTTGAACTGCCAA 3720  
Db 3661 CTGTTCTCAGGAGCTTCTATGAATCCCATGTTCCCATACCTTCCCAATTTGAACTGCCAA 3720  
Qy 3721 GAGAGTATCGAAACCGTTTCTTCTATGATGATGAGTGCAGGAATGGAGGCTTATCGAG 3780  
Db 3721 GAGAGTATCGAAACCGTTTCTTCTATGATGATGAGTGCAGGAATGGAGGCTTATCGAG 3780  
Qy 3781 ACAAAATGAAATTTGGACCCCTTGTGTACTAGCAACATTTGATTTACTATATCT 3840  
Db 3781 ACAAAATGAAATTTGGACCCCTTGTGTACTAGCAACATTTGATTTACTATATCT 3840

3841 QY CATTTTTCGTCGAGCAGTTAATTGCACTTAAGCGGAAGATATTTCCCGAAGAGGAGTAC 3900  
3841 Db CATTTTTCGTCGAGCAGTTAATTGCACTTAAGCGGAAGATATTTCCCGAAGAGGAGTAC 3900  
3901 QY ACAGAAGAGCTAGTCCCAATCGAATCAGAGTATAGAGATCTTCATTTGAAACCACTA 3950  
3901 Db ACAGAAGAGCTAGTCCCAATCGAATCAGAGTATAGAGATCTTCATTTGAAACCACTA 3950  
3961 QY COTCAGCATTTACTGAGCATTTTAAACCTCAGCTTCACAGAGATGTCTTTGTGTGTGAT 4020  
3961 Db COTCAGCATTTACTGAGCATTTTAAACCTCAGCTTCACAGAGATGTCTTTGTGTGTGAT 4020  
4021 QY GCTTAGCAGTTTGGCCGAGAGGAAATATCCAGTACCATGCTGTTTGTGGCATGAA 4080  
4021 Db GCTTAGCAGTTTGGCCGAGAGGAAATATCCAGTACCATGCTGTTTGTGGCATGAA 4080  
4081 QY TATAGCCCACTGACTAGGAATTAATTAACCAACCACTGAAACTTGTGTGCGAGCAGC 4140  
4081 Db TATAGCCCACTGACTAGGAATTAATTAACCAACCACTGAAACTTGTGTGCGAGCAGC 4140  
4141 QY TCTGAACCTGATTTTACTTTTAAAGAAATTTGCTCATGAGCCTGTCATCTCTTTTATAAAA 4200  
4141 Db TCTGAACCTGATTTTACTTTTAAAGAAATTTGCTCATGAGCCTGTCATCTCTTTTATAAAA 4200  
4201 QY GGCTCACTGACAGAGACAGCTGTTAAATTTCCACAGCAATCATTCGACAGCTAACTTTAT 4260  
4201 Db GGCTCACTGACAGAGACAGCTGTTAAATTTCCACAGCAATCATTCGACAGCTAACTTTAT 4260  
4261 QY TAGGAGAGCCTATGCCAGCTGGAGTGATTTGCTAAGAGGCTCCAGTCTTTGCAATCCAA 4320  
4261 Db TAGGAGAGCCTATGCCAGCTGGAGTGATTTGCTAAGAGGCTCCAGTCTTTGCAATCCAA 4320  
4321 QY AGCCTTTTCGTAAGCTTTGCACTTTTTCATTTCCATTTCCATTTTAAAGTAGTACTA 4380  
4321 Db AGCCTTTTCGTAAGCTTTGCACTTTTTCATTTCCATTTTCCATTTTAAAGTAGTACTA 4380  
4381 QY AGTTAACTAGTTATTTCTGCTTCTGAGTATAACGAATTTGGATGTCCTAAACCTATTTTA 4440  
4381 Db AGTTAACTAGTTATTTCTGCTTCTGAGTATAACGAATTTGGATGTCCTAAACCTATTTTA 4440  
4441 QY TAGATGTTATTTAAATATGAGCAATATCACCCTCTTATGCAATACCTAAATATGAG 4500  
4441 Db TAGATGTTATTTAAATATGAGCAATATCACCCTCTTATGCAATACCTAAATATGAG 4500  
4501 QY TTTTATTAATATTTAAGACTGTAATGTTCTTAAACCACTAACTCTAAGAGCTCAATG 4560  
4501 Db TTTTATTAATATTTAAGACTGTAATGTTCTTAAACCACTAACTCTAAGAGCTCAATG 4560  
4561 QY ATTGACATCTGAAATGCTTTGTAATTTATGACTTCAGCCCTTAAGAAATGCTATGATTC 4620  
4561 Db ATTGACATCTGAAATGCTTTGTAATTTATGACTTCAGCCCTTAAGAAATGCTATGATTC 4620  
4621 QY CGTCGAGGCTAAATTTCAACAGGCTAGAGTTAGTACTACTTACAGAGTGAATATGTTT 4680  
4621 Db CGTCGAGGCTAAATTTCAACAGGCTAGAGTTAGTACTACTTACAGAGTGAATATGTTT 4680  
4681 QY TGGAAATGACATATTTCAACAGAGAGTGCCTCATTTTGAATAATGATAGTCTGATGCA 4740  
4681 Db TGGAAATGACATATTTCAACAGAGAGTGCCTCATTTTGAATAATGATAGTCTGATGCA 4740  
4741 QY CTGGCACATTACAGTGGTCTGTTTAACTATCTGATATGTTATTCAGTACCTATCCT 4800  
4741 Db CTGGCACATTACAGTGGTCTGTTTAACTATCTGATATGTTATTCAGTACCTATCCT 4800  
4801 QY CTCAGTTGGTTTTCATAGAACAGAGGCGCAGCAAACTTTCTTTGTAAGAGGCTGGTTAGT 4860  
4801 Db CTCAGTTGGTTTTCATAGAACAGAGGCGCAGCAAACTTTCTTTGTAAGAGGCTGGTTAGT 4860  
4861 QY AAATTATTCGAGCCACCTGCTGCTTTGTCATACATTTCTTCTGCTGTTGTTAGTTGT 4920  
4861 Db AAATTATTCGAGCCACCTGCTGCTTTGTCATACATTTCTTCTGCTGTTGTTAGTTGT 4920  
4921 QY TTTTITTCACACACCTCTAAATGTTAAATGTTAAACCACTGTTTATGCTTGCAGCTGTACAAA 4980

4921 Db TTTTITTCACACACCTCTAAATGTTAAATGTTAAACCACTGTTTATGCTTGCAGCTGTACAAA 4980  
4981 QY ACTGCCACACAGCCAGATGTGACCCCTCAGGCCATCATTTGCCAATCACTGAGAAATTTT 5040  
4981 Db ACTGCCACACAGCCAGATGTGACCCCTCAGGCCATCATTTGCCAATCACTGAGAAATTTT 5040  
5041 QY TTGCTGTTGTTGTTGTTGTTGTTTGTGACAGAGTCTCTCTGTTGCCAGGCTGGAG 5100  
5041 Db TTGCTGTTGTTGTTGTTGTTGTTTGTGACAGAGTCTCTCTGTTGCCAGGCTGGAG 5100  
5101 QY TGCAGTGGGCCAATCTCAGCTCACTCCGCTCCCGGTTCCAGAGCAGTCTCTGTC 5160  
5101 Db TGCAGTGGGCCAATCTCAGCTCACTCCGCTCCCGGTTCCAGAGCAGTCTCTGTC 5160  
5161 QY TCAGCCTTCTGAGTAGTGGGACTA CAGGTGCGATGCCACACACACCCCTGCTAAATTTTGA 5220  
5161 Db TCAGCCTTCTGAGTAGTGGGACTA CAGGTGCGATGCCACACACACCCCTGCTAAATTTTGA 5220  
5221 QY TTTTATGATGAGACGGGGTTCCACCATATTTGTCAGGCTTATCTTTGAATCTCTGACCTC 5280  
5221 Db TTTTATGATGAGACGGGGTTCCACCATATTTGTCAGGCTTATCTTTGAATCTCTGACCTC 5280  
5281 QY AGGTGATCCACCTGCTCTGCTCCCAAAAGTGTGAGATTACAGGCAATAGCCAGTGCAC 5340  
5281 Db AGGTGATCCACCTGCTCTGCTCCCAAAAGTGTGAGATTACAGGCAATAGCCAGTGCAC 5340  
5341 QY CCAGCCGAGAAATAGTATTTTATGTTAAACCTTTGGCGTCTAGCCATATTTATG 5400  
5341 Db CCAGCCGAGAAATAGTATTTTATGTTAAACCTTTGGCGTCTAGCCATATTTATG 5400  
5401 QY TCATAATACAAATGGATTTGTGAAGAGCAGATTCCATGAGTAACCTCTGACAGGTAATTTAG 5460  
5401 Db TCATAATACAAATGGATTTGTGAAGAGCAGATTCCATGAGTAACCTCTGACAGGTAATTTAG 5460  
5461 QY ATCATGATCTCAACAATATTCCTCCAAATGGCATACATCTTTGTACAAAGAACTTGAA 5520  
5461 Db ATCATGATCTCAACAATATTCCTCCAAATGGCATACATCTTTGTACAAAGAACTTGAA 5520  
5521 QY ATGTAATACTGTTGTTGCTCTAAGAGTTGTATTTCAAAAACCTGAAATCTCATAA 5580  
5521 Db ATGTAATACTGTTGTTGCTCTAAGAGTTGTATTTCAAAAACCTGAAATCTCATAA 5580  
5581 QY AAGTTAAATTTTGAAAA 5597  
5581 Db AAGTTAAATTTTGAAAA 5597

## RESULT 5

AAD41199

ID AAD41199 standard; cdna; 4369 BP.

XX AAD41199;

XX AC AAD41199;

XX AC AAD41199;

XX DT 30-OCT-2002 (first entry)

XX Human nucleic acid-associated protein (NAAP-9) cdna.

Human; nucleic acid-associated protein; NAAP-9; neurological disorder; arteriosclerosis; cancer; cell proliferative disorder; atherosclerosis; lymphoma; epilepsy; Alzheimer's disease; developmental; anticonvulsant; autoimmune disorder; AIDS; allergy; anaemia; stroke; malaria; leishmania; gene therapy; neurotropic; neuroprotective; cerebroprotective; virucide; immunosuppressive; protozoacide; antimicrobial; gene; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 134..3904

XX /\*tag= a

XX /product= "Human NAAP-9"

XX PN WO200250279-A2.



QY 1532 TAATAAATCAGCCTCGAATGGGATGGTGGGATGGTCTGGAAACAGTGGAGGAGTCG 1591  
DB 1501 TAATAAATCAGCCTCGAATGGGATGGTGGGATGGTCTGGAAACAGTGGAGGAGTCG 1560  
QY 1592 CTATATTCAGAGGTGAGGTGAGTCTGGAGTATTCGAGTGGACACACCCCTCGCATTTGG 1651  
DB 1561 CTATATTCAGAGGTGAGGTGAGTCTGGAGTATTCGAGTGGACACACCCCTCGCATTTGG 1620  
QY 1652 TGGAGGAATAACAGTGTCTTACTGTAATCAGGGATGTGGGAATTCCTGCTCGCTGA 1711  
DB 1621 TGGAGGAATAACAGTGTCTTACTGTAATCAGGGATGTGGGAATTCCTGCTCGCTGA 1680  
QY 1712 TAAAGTCTGTGACCAAGCATGCAATGCTGTGCTGTGGTGTGATGTGCGCATGTGG 1771  
DB 1681 TAAAGTCTGTGACCAAGCATGCAATGCTGTGCTGTGGTGTGATGTGCGCATGTGG 1740  
QY 1772 GCAAGATCATTTTCATGATGTNTAAAGTGTGATCTCTCCCAACAGACTCACTATAT 1831  
DB 1741 GCAAGATCATTTTCATGATGTNTAAAGTGTGATCTCTCCCAACAGACTCACTATAT 1800  
QY 1832 TATTTCCAAAGGTGAATGCTGCTTATTTTTCAGCTTTTCGAGAAGTAGCCAAAGAGGAGT 1891  
DB 1801 TATTTCCAAAGGTGAATGCTGCTTATTTTCAGCTTTTCGAGAAGTAGCCAAAGAGGAGT 1860  
QY 1892 TGAAGGTGCTATAGTGAATCCAAATTAATTCGACATGCTTCTATTGCGCAACAGAGTGAA 1951  
DB 1861 TGAAGGTGCTATAGTGAATCCAAATTAATTCGACATGCTTCTATTGCGCAACAGAGTGAA 1920  
QY 1952 AACCATTCACCTCATAATGACAGTGGATGAATGCCACCAACACAAATTAATTTCTCAC 2011  
DB 1921 AACCATTCACCTCATAATGACAGTGGATGAATGCCACCAACACAAATTAATTTCTCAC 1980  
QY 2012 GTTTCAAAATAACAAACGATGAAGTTCAAAATGCAGATAACAGTGGAGGTGGACACAAG 2071  
DB 1981 GTTTCAAAATAACAAACGATGAAGTTCAAAATGCAGATAACAGTGGAGGTGGACACAAG 2040  
QY 2072 GGAGGACCAAACTGAATCTTACGGCCAGAGGTTACGAAATTTAGTTAGTCCCAT 2131  
DB 2041 GGAGGACCAAACTGAATCTTACGGCCAGAGGTTACGAAATTTAGTTAGTCCCAT 2100  
QY 2132 AACACTTTCTCAGAGGCGGAATCCTTTTGGAGATATTCCTCCAAAGAAACAGCTTCCC 2191  
DB 2101 AACACTTTCTCAGAGGCGGAATCCTTTTGGAGATATTCCTCCAAAGAAACAGCTTCCC 2160  
QY 2192 GAAATTTAAGACATGATGTTAACTCAACAGAGAGCCCAAGAGAGGTGAATAATCC 2251  
DB 2161 GAAATTTAAGACATGATGTTAACTCAACAGAGAGCCCAAGAGAGGTGAATAATCC 2220  
QY 2252 CTTGGTAAATATTTCACTCTCTCCAAAGACGCCAGTTGAGTCTCAATACCTTGGATTT 2311  
DB 2221 CTTGGTAAATATTTCACTCTCTCCAAAGACGCCAGTTGAGTCTCAATACCTTGGATTT 2280  
QY 2312 GCAATGGAAATGAGACATCACTTTGAAAGGATACAAATTTGTCCAAAGTCAGCTTGGCT 2371  
DB 2281 GCAATGGAAATGAGACATCACTTTGAAAGGATACAAATTTGTCCAAAGTCAGCTTGGCT 2340  
QY 2372 GAGATCATTTCTGATGAACTCACAGCATGCTAAATTAATAATCAAGCTATAATAACAGA 2431  
DB 2341 GAGATCATTTCTGATGAACTCACAGCATGCTAAATTAATAATCAAGCTATAATAACAGA 2400  
QY 2432 TGAACCAATGACAGTTTGGTGGCTCCACAGGAAACAGGTTCTATAAAGCATCTTGGCC 2491  
DB 2401 TGAACCAATGACAGTTTGGTGGCTCCACAGGAAACAGGTTCTATAAAGCATCTTGGCC 2460  
QY 2492 AAACAGCTTAGAGTGTCTGAAGATTCAGAGGTTGACTTTTCTGCGAGTGAAGTAAA 2551  
DB 2461 AAACAGCTTAGAGTGTCTGAAGATTCAGAGGTTGACTTTTCTGCGAGTGAAGTAAA 2520  
QY 2552 AGTGAATGGTCATGACCAAGGTTCAGAAATCCACCCCTGGACTTGGAGACCAAGCAAGATT 2611  
DB 2521 AGTGAATGGTCATGACCAAGGTTCAGAAATCCACCCCTGGACTTGGAGACCAAGCAAGATT 2580

QY 2612 TAGAGTGGAACTCACACCCCAAAAAACCATAGCGGAAATGTGACAAAAAGAACCCCC 2671  
DB 2581 TAGAGTGGAACTCACACCCCAAAAAACCATAGCGGAAATGTGACAAAAAGAACCCCC 2640  
QY 2672 ATCTCTGATTTTCCACTGTGAAAGCCAGATGACAAAGAAAAAATACACGGGAAAGA 2731  
DB 2641 ATCTCTGATTTTCCACTGTGAAAGCCAGATGACAAAGAAAAAATACACGGGAAAGA 2700  
QY 2732 AAAAGAGAACAGTAGAATGGAGGAAATGCTGAAAAATCACATAGGCGTTTACTGAAATGTT 2791  
DB 2701 AAAAGAGAACAGTAGAATGGAGGAAATGCTGAAAAATCACATAGGCGTTTACTGAAATGTT 2760  
QY 2792 ACTTGGAGAAAGCTGCGAGCANTTACAGATAGTTACTTTGGGCTTTTGGCCATGGGAGAA 2851  
DB 2761 ACTTGGAGAAAGCTGCGAGCANTTACAGATAGTTACTTTGGGCTTTTGGCCATGGGAGAA 2820  
QY 2852 AAAAAGATTTTCCAAAGATCTTCTCGACGAAGAGAGTCAATTCGAACACAAATTTGCATTA 2911  
DB 2821 AAAAAGATTTTCCAAAGATCTTCTCGACGAAGAGAGTCAATTCGAACACAAATTTGCATTA 2880  
QY 2912 CTTTCACTGATPAGCAAAAAATACTGGGAGGCACTTAAAGATACATTTTGCAGATTTCCCTCAG 2971  
DB 2881 CTTTCACTGATPAGCAAAAAATACTGGGAGGCACTTAAAGATACATTTTGCAGATTTCCCTCAG 2940  
QY 2972 ATATGTAATAATAATTTCTAAATAGCAAGTTTGGATTCACATCGCGGAAAGTCCCTGCTCA 3031  
DB 2941 ATATGTAATAATAATTTCTAAATAGCAAGTTTGGATTCACATCGCGGAAAGTCCCTGCTCA 3000  
QY 3032 CATGCTCTCATGATTCAGCCGATTTGTTATGCAAGAACTGCAAGATATGTTCCCTGAAGA 3091  
DB 3001 CATGCTCTCATGATTCAGCCGATTTGTTATGCAAGAACTGCAAGATATGTTCCCTGAAGA 3060  
QY 3092 ATTTGACAGACGTCTATTTTCAAAAGTGGCCATTTCTGAGGATATGCGATTTGCCCTTCTC 3151  
DB 3061 ATTTGACAGACGTCTATTTTCAAAAGTGGCCATTTCTGAGGATATGCGATTTGCCCTTCTC 3120  
QY 3152 TTATTTTATTTATCTCATGATGCTGAGTGCAGTGCACCTGAATATATCTCAAGTCTTTTATGA 3211  
DB 3121 TTATTTTATTTATCTCATGATGCTGAGTGCAGTGCACCTGAATATATCTCAAGTCTTTTATGA 3180  
QY 3212 AGTTGATPACAGATCAATCTGGTGTCTTGTCTGACAGAGAAATCCGAAACATGCTGCTACCAG 3271  
DB 3181 AGTTGATPACAGATCAATCTGGTGTCTTGTCTGACAGAGAAATCCGAAACATGCTGCTACCAG 3240  
QY 3272 AATTCAGAACCTGCGCTTAAAGTTGAGGATTTGACAGTCTGAGACATGCTGATATAA 3331  
DB 3241 AATTCAGAACCTGCGCTTAAAGTTGAGGATTTGACAGTCTGAGACATGCTGATATAA 3300  
QY 3332 TTGCTCAAAAAATCTTCTGCTGATATCAGCAGCTTAAATATAATATTTCCCAACTCAGGA 3391  
DB 3301 TTGCTCAAAAAATCTTCTGCTGATATCAGCAGCTTAAATATAATATTTCCCAACTCAGGA 3360  
QY 3392 ATCCTATCTATGATCCCAACCTGCGACCGGTCACTTAAAGTCTAGTAAACAACTGTAAACC 3451  
DB 3361 ATCCTATCTATGATCCCAACCTGCGACCGGTCACTTAAAGTCTAGTAAACAACTGTAAACC 3420  
QY 3452 AGTAACTGACAAAAATCCAAAGCATATAAGGACAAAAACAAATATAGGTTTCAAAATCAT 3511  
DB 3421 AGTAACTGACAAAAATCCAAAGCATATAAGGACAAAAACAAATATAGGTTTCAAAATCAT 3480  
QY 3512 GGGAGAGAAAGAAATCGCTTTTAAATTAAGTTCGTACCAACGTTTCTCATGTGTGGGCCA 3571  
DB 3481 GGGAGAGAAAGAAATCGCTTTTAAATTAAGTTCGTACCAACGTTTCTCATGTGTGGGCCA 3540  
QY 3572 GTTGGATGACATAAGAAAAAACCCCTAGGAAGTTTGTTCCTGAAATGACAACTTGCACCA 3631  
DB 3541 GTTGGATGACATAAGAAAAAACCCCTAGGAAGTTTGTTCCTGAAATGACAACTTGCACCA 3600  
QY 3632 CAATCATAAAGATGCTCAGACAGTGAAGCTGTTTCTCAGGAGCTTCTATGAATCCATGTT 3691  
DB 3601 CAATCATAAAGATGCTCAGACAGTGAAGCTGTTTCTCAGGAGCTTCTATGAATCCATGTT 3660  
QY 3692 CCCCATACCTTCCCAATTTGAACTGCCAAGAGAGTATCGAAACCGTTTCTTCCATCATGCA 3751









KW	N-acetylglucosamine-1-phosphodiester alpha-N-acetylglucosaminidase;
KW	enzyme replacement therapy; phosphorylated lysosomal hydrolase;
KW	lysosomal storage disease; gene; ss.
XX	
OS	Homo sapiens.
XX	
PN	US6537785-B1.
PN	
XX	
PD	25-MAR-2003.
XX	
XX	
PF	10-AUG-2000; 2000US-00636077.
XX	
PR	14-SEP-1999; 99US-0153831P.
XX	
PA	(GENZ-) GENZYME GLYCOBIOLOGY RES INST INC.
XX	
PI	Canfield WM;
XX	
DR	WPI; 2001-290356/30.
XX	
PT	Novel N-acetylglucosamine-1-phosphotransferase and N-acetylglucosamine-1-
PT	phosphodiester alpha-N-acetylglucosaminidase, useful for producing
PT	phosphorylated lysosomal hydrolase for treating lysosomal storage
PT	diseases.
XX	
XX	Example 9; Page 52-54; 62pp; English.
XX	
CC	The invention relates to a novel isolated human N-acetylglucosamine-1-
CC	phosphotransferase (GlcNAc-phosphotransferase) (I) and phosphodiester
CC	alpha-GlcNAcase (N-acetylglucosamine-1-phosphodiester alpha-N-
CC	Acetylglucosaminidase) (II). The protein of the invention has
CC	nephrotropic activity, and may be useful in enzyme replacement therapy. A
CC	protein of the invention (I), (II) is useful for preparing a
CC	phosphorylated lysosomal hydrolase. The phosphorylated hydrolase
CC	comprising a terminal mannose-6-phosphate, is useful for treating a
CC	patient suffering from a lysosomal storage disease. The present sequence
CC	is used in the exemplification of the invention
XX	
SQ	Sequence 3783 BP; 1198 A; 776 C; 835 G; 974 T; 0 U; 0 Other;
	Query Match 67.2%; Score 3761; DB 4; Length 3783;
	Best Local Similarity 99.7%; Pred. No. 0;
	Matches 3767; Conservative 0; Mismatches 10; Indels 0; Gaps 0
QY	165 ATGCTGTTCAAGCTCCTCGACAGACAAACCTATACCTGCTGCCACAGGTATGGGCTC 224
Db	
	7 ATGGGGTTCAAGCTCTTGACAGAGACAAACCTATACCTGCTGCCACAGGTATGGGCTC 66
QY	225 TAGCTGTGCTCTTCGGCGTGTGTGCACCATGCTCTCCGCTTCAGTTCGGAGAGTG 284
Db	
	67 TAGCTGTGCTCTTCGGCGTGTGTGCACCATGCTCTCCGCTTCAGTTCGGAGAGTG 126
QY	285 GTTCTGGAATGGAGCGAGATCAATACCATGTTTGTGTTTTCATTCTCTATAGACAAATATT 344
Db	
	127 GTTCTGGAATGGAGCGAGATCAATACCATGTTTGTGTTTTCATTCTCTATAGACAAATATT 186
QY	345 GCTGGAAAGTCCTTTCAGAACTGGCTTTTGCTGCCATGCCGATTCAGTGTGTTTACCC 404
Db	
	187 GCTGGAAAGTCCTTTCAGAACTGGCTTTTGCTGCCATGCCGATTCAGTGTGTTTACCC 246
QY	405 TGGGTGAATGGCAGACATCTTGAACACTAGAGGAACCTACAGCAGGTTCAGAGAACAGATG 464
Db	
	247 TGGGTGAATGGCAGACATCTTGAACACTAGAGGAACCTACAGCAGGTTCAGAGAACAGATG 306
QY	465 GAGGAGGAGCAGAAAGCAATAGAGAAATCTTTGGGAAAAACAACACGGAACTACTAAG 524
Db	
	307 GAGGAGGAGCAGAAAGCAATAGAGAAATCTTTGGGAAAAACAACACGGAACTACTAAG 366
QY	525 AAGAGTGAGACGCTTAGCTAGTGTGCTACACACTGCAATTAAGTGGCAATGCTTGTA 584
Db	
	367 AAGAGTGAGACGCTTAGCTAGTGTGCTACACACTGCAATTAAGTGGCAATGCTTGTC 426
QY	585 CTGGACCCAGCCCTCCACGCAACATCACCTGAGGACGTGCCATCTCTTTATCTCTCT 644

DB 1507 AGTCTCTTACTGTAATCAGGAGTGGCAATTCCTGCTCGCTGATAGATTCTGTGAC 1566  
QY 1725 CAAGCATGCAATGCTTGTGCTGTGGGTTTGATGCTGGCGACTGTGGGCAAGATCAATTT 1784  
DB 1567 CAAGCATGCAATGCTTGTGCTGTGGGTTTGATGCTGGCGACTGTGGGCAAGATCAATTT 1626  
QY 1785 CATGAATGTATAAAGTATCCTTCTCCAAACAGACTCACTATATTTATTCRAAAGGT 1844  
DB 1627 CATGAATGTATAAAGTATCCTTCTCCAAACAGACTCACTATATTTATTCRAAAGGT 1686  
QY 1845 GAATGCTGCTTATTTAGCTTTGCGAAGTAGCCAAAAGAGAGAGTTGAAGTGGCTAT 1904  
DB 1687 GAATGCTGCTTATTTAGCTTTGCGAAGTAGCCAAAAGAGAGAGTTGAAGTGGCTAT 1746  
QY 1905 AGTGACAATCAATAATTCGACATGCTTCTATTGCCAAACAGTGGAAACCATCCACCTC 1964  
DB 1747 AGTGACAATCCAAATTCGACATGCTTCTATTGCCAAACAGTGGAAACCATCCACCTC 1806  
QY 1965 ATAATGCACAGTGAATGAATGCCACCAATACATTTTAAATCTCAGCTTTCAAAATACA 2024  
DB 1807 ATAATGCACAGTGAATGAATGCCACCAATACATTTTAAATCTCAGCTTTCAAAATACA 1866  
QY 2025 AACCATGAGAGTTCAAAATGCAGATAACAGTGGAGTGACACAGGAGGAGGACCAAA 2084  
DB 1867 AACCATGAGAGTTCAAAATGCAGATAACAGTGGAGTGACACAGGAGGAGGACCAAA 1926  
QY 2085 CTGAATCTCAGGCCCAAGAGGTTAGCAAAATTTAGTTAGTCTCCATAACACTTTTCCA 2144  
DB 1927 CTGAATCTCAGGCCCAAGAGGTTAGCAAAATTTAGTTAGTCTCCATAACACTTTTCCA 1986  
QY 2145 GAGCGGGAATTCCTTTTGGAGTATTCCTCAAGAAAAGCTTCCCGAAGTTTAAGAGA 2204  
DB 1987 GAGCGGGAATTCCTTTTGGAGTATTCCTCAAGAAAAGCTTCCCGAAGTTTAAGAGA 2046  
QY 2205 CATGATGTTAACTCAAGAGAGAGCCAGAGAGAGTGAATTCCTCGTGAATATTT 2264  
DB 2047 CATGATGTTAACTCAAGAGAGAGCCAGAGAGAGTGAATTCCTCGTGAATATTT 2106  
QY 2265 TCACCTCTTCCAAAGAGCCAGCTTGAATCTCAATACCTTGGATTTGCAACTGGAACAT 2324  
DB 2107 TCACCTCTTCCAAAGAGCCAGCTTGAATCTCAATACCTTGGATTTGCAACTGGAACAT 2166  
QY 2325 GGAGACATCACTTTGAAGAGTACAATTTGTCGAAGTCAAGTCAAGTCTGAGATCATTTCTG 2384  
DB 2167 GGAGACATCACTTTGAAGAGTACAATTTGTCGAAGTCAAGTCAAGTCTGAGATCATTTCTG 2226  
QY 2385 ATGAACCTCAGAGCTGCTAAATATAAATAAATCAAGCTATAATAACAGATGAACAAATGAC 2444  
DB 2227 ATGAACCTCAGAGCTGCTAAATATAAATAAATCAAGCTATAATAACAGATGAACAAATGAC 2286  
QY 2445 AGTTTGTGGCTCCACAGGAAAACAGGTTCAATAAAGCATCTGCCAAGACGTTAGGA 2504  
DB 2287 AGTTTGTGGCTCCACAGGAAAACAGGTTCAATAAAGCATCTGCCAAGACGTTAGGA 2346  
QY 2505 GTGTCTGAAAGATTGCGAGAGTTGACATTTTCTCGAGTGAATGTAATAAGTGAATGCTCAT 2564  
DB 2347 GTGTCTGAAAGATTGCGAGAGTTGACATTTTCTCGAGTGAATGTAATAAGTGAATGCTCAT 2406  
QY 2565 GACAGGCTGAGATCCACCTCGAATTGGAGACCAAGAGATTTAGAGTGGAACT 2624  
DB 2407 GACAGGCTGAGATCCACCTCGAATTGGAGACCAAGAGATTTAGAGTGGAACT 2466  
QY 2625 CACACCCAAAACCATAGGCGGAAATGTGACAAAAGAAAGCCCATCTCTGATTGTT 2684  
DB 2467 CACACCCAAAACCATAGGCGGAAATGTGACAAAAGAAAGCCCATCTCTGATTGTT 2526  
QY 2685 CCACCTGGAAGCCAGATGACAAAAGAAAGAAATCA CAGGAAAAGAAAGAAAGACAGT 2744  
DB 2527 CCACCTGGAAGCCAGATGACAAAAGAAAGAAATCA CAGGAAAAGAAAGAAAGACAGT 2586  
QY 2745 AGRATGGAGGAAATGCTGAAATACATAGGCGCTTACTGAAGTGTACTTTGGAAGAAAG 2804  
DB 2587 AGRATGGAGGAAATGCTGAAATACATAGGCGCTTACTGAAGTGTACTTTGGAAGAAAG 2646

QY 2805 CTGCAGCAATTACAGATAGATTACTTGGCTTTTTCCTCATGGGAGAAAAAAGTATTTTC 2864  
DB 2647 CTGCAGCAATTACAGATAGATTACTTGGCTTTTTCCTCATGGGAGAAAAAAGTATTTTC 2706  
QY 2865 CAAGATCTTCTCGACGAAGAAGAGTCAATTGAAGACACAATTTGGCATCTCTCAGTATAGC 2924  
DB 2707 CTAGATCTTCTCGACGAAGAAGAGTCAATTGAAGACACAATTTGGCATCTCTCAGTATAGC 2766  
QY 2925 AAAATATCTGGAGGCACTAAAGATACATTTGCAGATTCCTCTCAGATATGTAATAAA 2984  
DB 2767 AAGATATCTGGAGGCACTAAAGATACATTTGCAGATTCCTCTCAGATATGTAATAAA 2826  
QY 2985 ATTCTAAATAGCAAGTTTGGATTCAATPCGCGGAAGTCCCTGCTCAATGCCTCACATG 3044  
DB 2827 ATTCTAAATAGCAAGTTTGGATTCAATPCGCGGAAGTCCCTGCTCAATGCCTCACATG 2886  
QY 3045 ATTGACCGGATGTTATGCAAGAACTGCAAGATATGTTCCCTGAAGATTTGCAAGACG 3104  
DB 2887 ATTGACCGGATGTTATGCAAGAACTGCAAGATATGTTCCCTGAAGATTTGCAAGACG 2946  
QY 3105 TCATTTTCAAAAGTGGCCATTCTGAGGATATGAGATTTGCTTCTCTTATTTTATAT 3164  
DB 2947 TCATTTTCAAAAGTGGCCATTCTGAGGATATGAGATTTGCTTCTCTTATTTTATAT 3006  
QY 3165 CTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGAT 3224  
DB 3007 CTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGAT 3066  
QY 3225 CAATCTGTGTCTGTCTGACAGAGAAATCCGAACACTGGCTACAGAAATTCAGAACTG 3284  
DB 3067 CAATCTGTGTCTGTCTGACAGAGAAATCCGAACACTGGCTACAGAAATTCAGAACTG 3126  
QY 3285 CGTTTAAAGTTTGCAGGATTTGACAGGCTTGAAACACATGCTTAATAAATTTGCTCAAAATG 3344  
DB 3127 CGTTTAAAGTTTGCAGGATTTGACAGGCTTGAAACACATGCTTAATAAATTTGCTCAAAATG 3186  
QY 3345 CTTCTCTGTGATATCACCGAGCTAAATAATTTCCCAACTCAGGATCTCTACTATGAT 3404  
DB 3187 CTTCTCTGTGATATCACCGAGCTAAATAATTTCCCAACTCAGGATCTCTACTATGAT 3246  
QY 3405 CCGAATCTGCCACCGCTCACTAAAGTCTAGTAAACAACTGTAACCCAGTAACTGACAAA 3464  
DB 3247 CCGAATCTGCCACCGCTCACTAAAGTCTAGTAAACAACTGTAACCCAGTAACTGACAAA 3306  
QY 3465 ATCCCAAGAGCATATAAGGACAAAACAATATAGTTTGAATCATCGGAGAGAGAA 3524  
DB 3307 ATCCCAAGAGCATATAAGGACAAAACAATATAGTTTGAATCATCGGAGAGAGAA 3366  
QY 3525 ATCGCTTTTAAATGATTCGTACCAACGTTTCTCATGTGGTGGCCAGTTGGATGACATA 3584  
DB 3367 ATCGCTTTTAAATGATTCGTACCAACGTTTCTCATGTGGTGGCCAGTTGGATGACATA 3426  
QY 3585 AGAAAAACCCCTAGGAAGTTTGTTCCTGATGAACCAATTCATGACCAATCATAAAGAT 3644  
DB 3427 AGAAAAACCCCTAGGAAGTTTGTTCCTGATGAACCAATTCATGACCAATCATAAAGAT 3486  
QY 3645 GCTCAGACAGTGAAGGCTGTTCTCAGGACATCTTATGAATCCATGTTCCCATACCTTCC 3704  
DB 3487 GCTCAGACAGTGAAGGCTGTTCTCAGGACATCTTATGAATCCATGTTCCCATACCTTCC 3546  
QY 3705 CAATTTGAATCTGCCAAGAGATATCGAAACCCGTTTCTTCATATGATGAGCTGCGAGAA 3764  
DB 3547 CAATTTGAATCTGCCAAGAGATATCGAAACCCGTTTCTTCATATGATGAGCTGCGAGAA 3606  
QY 3765 TGGAGGCTTATCGACACAAATTTGAAGTTTGGCCCACTTGTACTAGCAACATTTGAT 3824  
DB 3607 TGGAGGCTTATCGACACAAATTTGAAGTTTGGCCCACTTGTACTAGCAACATTTGAT 3666  
QY 3825 ATGTTTACTATATTTCTCATTTTGTCTGAGCAGTAAATTTGCACTTAAAGCGGAAGATTTT 3884  
DB 3667 ATGTTTACTATATTTCTCATTTTGTCTGAGCAGTAAATTTGCACTTAAAGCGGAAGATTTT 3726

QY 3885 CCCAAGGAGGATACACAAAGAGCTAGTCCCAATCGAATCAGAGTATAGAAGTC 3941  
DB 3727 CCCAAGGAGGATACACAAAGAGCTAGTCCCAATCGAATCAGAGTATAGAAGTC 3783

RESULT 8

ACC81011  
ID ACC81011 standard; cDNA; 3621 BP.  
XX ACC81011;  
AC ACC81011;  
XX  
DT 01-AUG-2003 (first entry)  
XX  
DE Human GlcNAc-phosphotransferase related cDNA SEQ ID NO: 21.  
XX  
KW Human; N-acetylglucosamine-1-phosphotransferase; nephrotropic;  
KW GlcNAc-phosphotransferase; phosphodiester alpha-GlcNAcase;  
KW N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase;  
KW enzyme replacement therapy; phosphorylated lysosomal hydrolase;  
KW lysosomal storage disease; gene; ss.  
XX  
OS Homo sapiens.  
XX  
FN US6537785-B1.  
XX  
PD 25-MAR-2003.  
XX  
PF 10-AUG-2000; 2000US-00636077.  
XX  
PR 14-SEP-1999; 99US-0153831P.  
XX  
PA (GENZ-) GENZYME GLYCOBIOLOGY RES INST INC.  
XX  
PI Canfield WM;  
XX  
PS WPI; 2001-290356/30.  
XX

XX Novel N-acetylglucosamine-1-phosphotransferase and N-acetylglucosamine-1-  
PT phosphodiester alpha-N-Acetylglucosaminidase, useful for producing  
PT phosphorylated lysosomal hydrolase for treating lysosomal storage  
PT diseases.  
XX  
PS Example 10; Page 54-55; 62pp; English.  
XX  
CC The invention relates to a novel isolated human N-acetylglucosamine-1-  
CC phosphotransferase (GlcNAc-phosphotransferase) (I) and phosphodiester  
CC alpha-GlcNAcase (N-acetylglucosamine-1-phosphodiester alpha-N-  
CC Acetylglucosaminidase) (II). The protein of the invention has  
CC nephrotropic activity, and may be useful in enzyme replacement therapy. A  
CC protein of the invention (I), (II) is useful for preparing a  
CC phosphorylated lysosomal hydrolase. The phosphorylated hydrolase  
CC comprising a terminal mannose-6-phosphate, is useful for treating a  
CC patient suffering from a lysosomal storage disease. The present sequence  
CC is used in the exemplification of the invention  
XX  
SQ Sequence 3621 BP; 1156 A; 743 C; 802 G; 920 T; 0 U; 0 Other;

Query Match 62.3%; Score 3485.4; DB 4; Length 3621;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 3492; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 297 ACCGAGATCATACCATGTTTGTTCGATTCCTATACAGACATATGCTGGAAGTCC 356  
DB 115 ACCGAGATCATACCATGTTTGTTCGATTCCTATACAGACATATGCTGGAAGTCC 174  
QY 357 TTTCAGATCGGCTTTGTCTGCCATGCCGATGACGTTGTTTACACCTGGGTGAATGCC 416  
DB 175 TTTCAGATCGGCTTTGTCTGCCATGCCGATGACGTTGTTTACACCTGGGTGAATGCC 234  
QY 417 ACAGATCTTGAACACTGAGGAAGTACAGAGGTACAGACACAGATGAGGAGGACGAG 476  
DB 235 ACAGATCTTGAACACTGAGGAAGTACAGAGGTACAGACACAGATGAGGAGGAGGAG 294

QY 477 AAGCAATGAGAGAAATCCTTGGGAAACACACGGAACCTACTAAGAGAGTGAGAAG 536  
DB 295 AAGCAATGAGAGAAATCCTTGGGAAACACACGGAACCTACTAAGAGAGTGAGAAG 354  
QY 537 CAGTTAGAGTGTGCTTAACACACACTGCATTAAAGTGCCCAATGCTTCTCTGGACCCAGCC 596  
DB 355 CAGTTAGAGTGTGCTTAACACACACTGCATTAAAGTGCCCAATGCTTCTCTGGACCCAGCC 414  
QY 597 CTGCCAGCCACATCAACCTCTGAAGGACCTGCCATCTCTTTATCTCTCTCTCTCTCTG 656  
DB 415 CTGCCAGCCACATCAACCTCTGAAGGACCTGCCATCTCTTTATCTCTCTCTCTCTCTG 474  
QY 657 AGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCTCAGTTGTGTT 716  
DB 475 AGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCTCAGTTGTGTT 534  
QY 717 TTTGACAGTACTAAGGATGTTGAAGATGCCACTCTGGAATGCTCTTAAAGAAATAGCAGA 776  
DB 535 TTTGACAGTACTAAGGATGTTGAAGATGCCACTCTGGAATGCTCTTAAAGAAATAGCAGA 594  
QY 777 CAGACAGTATGAGGGGTACTTTGACACAGATTAAGAAAGTCCCTCGATTAGTGCTAATG 836  
DB 595 CAGACAGTATGAGGGGTACTTTGACACAGATTAAGAAAGTCCCTCGATTAGTGCTAATG 654  
QY 837 CAAGATTTGGCTTTCTGAGTGGATTTCCACCAACATTCAGGAAACCAATCAACTAAAA 896  
DB 655 CAAGATTTGGCTTTCTGAGTGGATTTCCACCAACATTCAGGAAACCAATCAACTAAAA 714  
QY 897 ACAAATTCGAGAAATCTTTCTCTAAGTCAAACTGTTGAGTGTGTTTATTCAGAGGCC 956  
DB 715 ACAAATTCGAGAAATCTTTCTCTAAGTCAAACTGTTGAGTGTGTTTATTCAGAGGCC 774  
QY 957 AGTGTAGCGCTTCTAAACCTGAATAACCCCAAGGATTTTCAAGAAATGAATGAACAACT 1016  
DB 775 AGTGTAGCGCTTCTAAACCTGAATAACCCCAAGGATTTTCAAGAAATGAATGAACAACT 834  
QY 1017 AAGAAGACATGACCATGATGGAAGAACTGACCATAGTCTCTCATATTTATTTATGG 1076  
DB 835 AAGAAGACATGACCATGATGGAAGAACTGACCATAGTCTCTCATATTTATTTATGG 894  
QY 1077 GATCTGAGGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCCAGTGTGTTTGA 1136  
DB 895 GATCTGAGGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCCAGTGTGTTTGA 954  
QY 1137 GATAACGAAGAACTGAGGTACTCATTTGCCATCTATCGAGAGGATGACCATGGGTCCG 1196  
DB 955 GATAACGAAGAACTGAGGTACTCATTTGCCATCTATCGAGAGGATGACCATGGGTCCG 1014  
QY 1197 AATATTTTCAATGTCACCAACGGCGAGATTCCTCTGCTGAACCTTGACAACTCTCGA 1256  
DB 1015 AATATTTTCAATGTCACCAACGGCGAGATTCCTCTGCTGAACCTTGACAACTCTCGA 1074  
QY 1257 GTGACAAATAGTAAACACACAGGATGTTTTTCGAAATTTGAGCCACTTGCTTACTTACT 1316  
DB 1075 GTGACAAATAGTAAACACACAGGATGTTTTTCGAAATTTGAGCCACTTGCTTACTTACT 1134  
QY 1317 TCACCTGCTATTGAAAGTCACATTCATCGATCGAAGGCTGTCCGAGAAAGTTTATTAC 1376  
DB 1135 TCACCTGCTATTGAAAGTCACATTCATCGATCGAAGGCTGTCCGAGAAAGTTTATTAC 1194  
QY 1377 CTAATATGATGATGTCATGTTTGGGAAGGATGTCTGGCCAGATGATTTTACAGTCACTCC 1436  
DB 1195 CTAATATGATGATGTCATGTTTGGGAAGGATGTCTGGCCAGATGATTTTACAGTCACTCC 1254  
QY 1437 AAGGCCGAGAGGTTTATTTGACATGGCTGTGCCAAACTGTGCCAGGCTGCCAGGT 1496  
DB 1255 AAGGCCGAGAGGTTTATTTGACATGGCTGTGCCAAACTGTGCCAGGCTGCCAGGT 1314  
QY 1497 TCCTGATTAAGGATGGCTATTGTGACAAAGGCTTGTAAATTCAGCTTCGATGGGAT 1556  
DB 1315 TCCTGATTAAGGATGGCTATTGTGACAAAGGCTTGTAAATTCAGCTTCGATGGGAT 1374  
QY 1557 GGTGGGATGCTCTGGAAACAGTGGAGGAGTCTGCTATATTGACAGAGGTGGAGTACT 1616



Db 1375 GGTGGGATTTGCTCTGGAAACAGTGGAGGAGTGCCTATATTGACGAGGTGGAGTACT 1434  
QY 1617 GGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTTTAC 1676  
Db 1435 GGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTTTAC 1494  
QY 1677 TGTAAACAGGATGTCGAATTTCTGGCTGCTGCTGATTAAGTTCGTGACCAAGCATCAAT 1736  
Db 1495 TGTAAACAGGATGTCGAATTTCTGGCTGCTGCTGATTAAGTTCGTGACCAAGCATCAAT 1554  
QY 1737 GTCTTTGCTCTGGGTTTGTATGCTGGGACTGTGGGCAAGATCAATTTTCATGAATTTGAT 1796  
Db 1555 GFTTTGCTCTGGGTTTGTATGCTGGGACTGTGGGCAAGATCAATTTTCATGAATTTGAT 1614  
QY 1797 AAAGTGATCTTCTCCCAACACAGACTCACTATATATTTCAAAAGTGAATGCTGCTCCT 1856  
Db 1615 AAAGTGATCTTCTCCCAACACAGACTCACTATATATTTCAAAAGTGAATGCTGCTCCT 1674  
QY 1857 TATTTTCAGCTTTGAGAAAGTAGCCAAAAGAGAGCTTGAAGTGCCTATAGTGACAATCCA 1916  
Db 1675 TATTTTCAGCTTTGAGAAAGTAGCCAAAAGAGAGCTTGAAGTGCCTATAGTGACAATCCA 1734  
QY 1917 ATAAATTCAGATGCTTTTATTTGCCAAACAGTGGGAAAACCACTCCATTAATAGCAGT 1976  
Db 1735 ATAAATTCAGATGCTTTTATTTGCCAAACAGTGGGAAAACCACTCCATTAATAGCAGT 1794  
QY 1977 GGAATGAATGCCACACACATATTTTAACTCTCAGCTTTCAAAATACAAACAGTGAAGAG 2036  
Db 1795 GGAATGAATGCCACACATATTTTAACTCTCAGCTTTCAAAATACAAACAGTGAAGAG 1854  
QY 2037 TTCAAAATGACAGATAACAGTGGAGTGGACACAAGGAGGAGCAAAACCTGAATTTACG 2096  
Db 1855 TTCAAAATGACAGATAACAGTGGAGTGGACACAAGGAGGAGCAAAACCTGAATTTACG 1914  
QY 2097 GCCCAGAGGCTTACGAAAATTTAGTTAGTCCCATACACTTCTTCCAGAGGCGGAATC 2156  
Db 1915 GCCCAGAGGCTTACGAAAATTTAGTTAGTCCCATACACTTCTTCCAGAGGCGGAATC 1974  
QY 2157 CTTTTTGAAGGATATTTCCAAAGAAAACCGCTTCCCGAAGTTTAAAGACATGATGTTAAC 2216  
Db 1975 CTTTTTGAAGGATATTTCCAAAGAAAACCGCTTCCCGAAGTTTAAAGACATGATGTTAAC 2034  
QY 2217 TCACAAGGAGAGCCCAAGAGAGGTGAAAATTTCCCTCGTGAATTAATTTCACTCTCTTCCA 2276  
Db 2035 TCACAAGGAGAGCCCAAGAGAGGTGAAAATTTCCCTCGTGAATTAATTTCACTCTCTTCCA 2094  
QY 2277 AAAGACGCCAGTTGAGTCTCAATACCTTGGATTGGAACTGGAACTGGAGACATCACT 2336  
Db 2095 AAAGACGCCAGTTGAGTCTCAATACCTTGGATTGGAACTGGAACTGGAGACATCACT 2154  
QY 2337 TTGAAAGATACAAATTTGTCCAACTCAGCTTGTGAGATCAATTTCTGATGAATCTCACAG 2396  
Db 2155 TTGAAAGATACAAATTTGTCCAACTCAGCTTGTGAGATCAATTTCTGATGAATCTCACAG 2214  
QY 2397 CATGCTAAAATAAAAATCAAGCTATATAACAGATGAACAAATGACAGTTTGGTGGCT 2456  
Db 2215 CATGCTAAAATAAAAATCAAGCTATATAACAGATGAACAAATGACAGTTTGGTGGCT 2274  
QY 2457 CCACAGGAAAAACAGGTTTCATAAAGCATCTTGCAAAACAGCTTTAGAGGTGCTCGAAAGA 2516  
Db 2275 CCACAGGAAAAACAGGTTTCATAAAGCATCTTGCAAAACAGCTTTAGAGGTGCTCGAAAGA 2334  
QY 2517 TTGACAGAGTTTGAATTTCTTCCAGTGAAGTGAAGTGAATGCTCATGACCAAGGCTCAG 2576  
Db 2335 TTGACAGAGTTTGAATTTCTTCCAGTGAAGTGAAGTGAATGCTCATGACCAAGGCTCAG 2394  
QY 2577 AATCCACCCCTGGACTTGGAGACCAACAGCAAGATTTAGAGTGGAAACTCACACCCAAAAA 2636  
Db 2395 AATCCACCCCTGGACTTGGAGACCAACAGCAAGATTTAGAGTGGAAACTCACACCCAAAAA 2454  
QY 2637 ACCATAGCGGAAATGTGACAAAGAAAAGCCGCCATCTCTGATTTTCACTGGAAAGC 2696

Db 2455 ACCATAGCGGAAATGTGACAAAGAAAAGCCCCATCTCTGATTTGTTCTCACTGGAAAGC 2514  
QY 2697 CAGATGACAAAAGAAAAGAAAATCAACAGGAAAAGAAAAGAGAACAGTAGAATGAGAGAA 2756  
Db 2515 CAGATGACAAAAGAAAAGAAAATCAACAGGAAAAGAAAAGAGAACAGTAGAATGAGAGAA 2574  
QY 2757 AATGCTGAAAATCACATAGGCGTTACTGAAGTGTTCATTTGGAAGAAAGCTCAGCATTAC 2816  
Db 2575 AATGCTGAAAATCACATAGGCGTTACTGAAGTGTTCATTTGGAAGAAAGCTCAGCATTAC 2634  
QY 2817 ACAGATAGTTACTTGGGCTTTTGGCATGGGAGAAAAAAGTATTTTCCAAAGATCTTCTC 2876  
Db 2635 ACAGATAGTTACTTGGGCTTTTGGCATGGGAGAAAAAAGTATTTTCTAGATCTTCTC 2694  
QY 2877 GACGAGAGAGTCAATTTGAAGACACAATTTGACATACTTCACTGATAGCAAAAATACTGGG 2936  
Db 2695 GACGAGAGAGTCAATTTGAAGACACAATTTGACATACTTCACTGATAGCAAAAATACTGGG 2754  
QY 2937 AGGCAACTAAAAGATACATTTGCGAGATTCCTTCAGATATGTAATAAATTTCTAAATAGC 2996  
Db 2755 AGGCAACTAAAAGATACATTTGCGAGATTCCTTCAGATATGTAATAAATTTCTAAATAGC 2814  
QY 2997 AAGTTTGGATTCACATCGGAAAGTCCCTGCTCAGATGCCTCACATGATGACCGGAT 3056  
Db 2815 AAGTTTGGATTCACATCGGAAAGTCCCTGCTCAGATGCCTCACATGATGACCGGAT 2874  
QY 3057 GTTATGCAAGAACTGCAAGATATGTTCCCTGAAAGAAATTTGAACAGACGTCATTTCAAAA 3116  
Db 2875 GTTATGCAAGAACTGCAAGATATGTTCCCTGAAAGAAATTTGAACAGACGTCATTTCAAAA 2934  
QY 3117 GTGCGCCATTTGAGGATATGCGAGTATGCGGTTGCTTCTTATTTTATTTATCTCATGAGTGA 3176  
Db 2935 GTGCGCCATTTGAGGATATGCGAGTATGCGGTTGCTTCTTATTTTATTTATCTCATGAGTGA 2994  
QY 3177 GTGCGACCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTGTG 3236  
Db 2995 GTGCGACCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTGTG 3054  
QY 3237 TTGCTCTGACAGAGAAATCCGAACACTGCTACCAAGATTCACGAATTCGCCGTTAAGTTTG 3296  
Db 3055 TTGCTCTGACAGAGAAATCCGAACACTGCTACCAAGATTCACGAATTCGCCGTTAAGTTTG 3114  
QY 3297 CAGGATTTGACAGGCTCTGGAACACATGCTTAATAATTTGCTCABAAAATGCTTCTGCTGAT 3356  
Db 3115 CAGGATTTGACAGGCTCTGGAACACATGCTTAATAATTTGCTCABAAAATGCTTCTGCTGAT 3174  
QY 3357 ATCAGCGAGCTAAAATAATTTCCACCAACTCAGGAATCTTACTATGATCCCAACCTGCCA 3416  
Db 3175 ATCAGCGAGCTAAAATAATTTCCACCAACTCAGGAATCTTACTATGATCCCAACCTGCCA 3234  
QY 3417 CCGGTCACCTAAAGTCTAGTAAACAACTGTAAACCACTGAACTGACAAAATCCACAGACGA 3476  
Db 3235 CCGGTCACCTAAAGTCTAGTAAACAACTGTAAACCACTGAACTGACAAAATCCACAGACGA 3294  
QY 3477 TATAAGGACAAAACAAATATAGGTTTGAATTCATGGGAGAAAGAAATCGCTTTTAAA 3536  
Db 3295 TATAAGGACAAAACAAATATAGGTTTGAATTCATGGGAGAAAGAAATCGCTTTTAAA 3354  
QY 3537 ATGATTCGTACCAAGTTTCTCATGTGGTGGCAGTTGGATGACATGAAGAAAAAACCTC 3596  
Db 3355 ATGATTCGTACCAAGTTTCTCATGTGGTGGCAGTTGGATGACATGAAGAAAAAACCTC 3414  
QY 3597 AGGAAGTTTGTGCTGAATGACAACTTCCACCACTGAACTGAACTGAACTGAACTGAACTG 3656  
Db 3415 AGGAAGTTTGTGCTGAATGACAACTTCCACCACTGAACTGAACTGAACTGAACTGAACTG 3474  
QY 3657 AAGGCTGTTCTCAGGACTTCTATGAATTCATGTTTCCCATACCTTCCCAATTTGAACTG 3716  
Db 3475 AAGGCTGTTCTCAGGACTTCTATGAATTCATGTTTCCCATACCTTCCCAATTTGAACTG 3534  
QY 3717 CCAGAGAGTATCCGAACCGGTTTCTTCCATATGATGCTGAGGAGTATGAGGAGCTTAT 3776  
Db 3535 CCAGAGAGTATCCGAACCGGTTTCTTCCATATGATGCTGAGGAGTATGAGGAGCTTAT 3594





1557 QY GGTGGGATTTGCTCTGGAAACAGTGGAGGAGTGCCTATATTGACGAGGTGGAGGTACT 1616  
1363 Db GGTGGGATTTGCTCTGGAAACAGTGGAGGAGTGCCTATATTGACGAGGTGGAGGTACT 1422  
1617 QY GGGAGTATGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAAGTGTCTTTAC 1676  
1423 Db GGGAGTATGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAAGTGTCTTTAC 1482  
1677 QY TGTAATCAGGAGTGGCAATTCCTGGCTCCCTGATAGTTCTGACCAAGCATGCAAT 1736  
1483 Db TGTAATCAGGAGTGGCAATTCCTGGCTCCCTGATAGTTCTGACCAAGCATGCAAT 1542  
1737 QY GTCTTGCTCTGGGTTTGATGCTGGGACTGTGGGCAAGATCATTTTCATGAATGTAT 1796  
1543 Db GTCTTGCTCTGGGTTTGATGCTGGGACTGTGGGCAAGATCATTTTCATGAATGTAT 1602  
1797 QY AAGTGATCCTTCTCCCAACACAGCTCACTATATATTATTCAAAAGGTGAATGCTGCCT 1856  
1603 Db AAGTGATCCTTCTCCCAACACAGCTCACTATATATTATTCAAAAGGTGAATGCTGCCT 1662  
1857 QY TATTTTCACTTTGACAGTGTAGCCAAAGAGGAGTGAAGGTGCTATAGTGACAAATCCA 1916  
1663 Db TATTTTCACTTTGACAGTGTAGCCAAAGAGGAGTGAAGGTGCTATAGTGACAAATCCA 1722  
1917 QY ATAATTCAGATGCTTTCTATTGCCAAACAGTGTGAAAACCAATCCACCTCATATAATGACAGT 1976  
1723 Db ATAATTCAGATGCTTTCTATTGCCAAACAGTGTGAAAACCAATCCACCTCATATAATGACAGT 1782  
1977 QY GGAATGAATGCCACACAAATCACTTTTAATCTCAGTTTCAAAATACAAACGATGAAGAG 2036  
1783 Db GGAATGAATGCCACACAAATCACTTTTAATCTCAGTTTCAAAATACAAACGATGAAGAG 1842  
2037 QY TTCAAAATGACAGTAACAGTGGAGTGGACACAAGGAGGAGCAAAACCTGAATTTACG 2096  
1843 Db TTCAAAATGACAGTAACAGTGGAGTGGACACAAGGAGGAGCAAAACCTGAATTTACG 1902  
2097 QY GCCCAGAGGGTTACGAAATTTAGTTAGTCCCATACACTTTTCCAGAGGGCGGAATC 2156  
1903 Db GCCCAGAGGGTTACGAAATTTAGTTAGTCCCATACACTTTTCCAGAGGGCGGAATC 1962  
2157 QY CTTTGTGAGGATATTTCCAAAGAAAACCGTTCCTCCGAAAGTTTAAGACACATGATTTAAC 2216  
1963 Db CTTTGTGAGGATATTTCCAAAGAAAACCGTTCCTCCGAAAGTTTAAGACACATGATTTAAC 2022  
2217 QY TCACAGGAGAGCCAGGAGAGGTGAATTTCCCTGTTAAATTTTCACTCTCTTCCA 2276  
2023 Db TCACAGGAGAGCCAGGAGAGGTGAATTTCCCTGTTAAATTTTCACTCTCTTCCA 2082  
2277 QY AAAGACGCCAGTTGAGTCTCAATACCTTGGATTGCAACTGGAAACATGGAGACATCACT 2336  
2083 Db AAAGACGCCAGTTGAGTCTCAATACCTTGGATTGCAACTGGAAACATGGAGACATCACT 2142  
2337 QY TTGAAGAGATACAAATTTGTCAGATGACCTTGTCTGAGATCAATTTCTGATGAATCTCACAG 2396  
2143 Db TTGAAGAGATACAAATTTGTCAGATGACCTTGTCTGAGATCAATTTCTGATGAATCTCACAG 2202  
2397 QY CATGCTAAATTAATAATCAAGCTATATAACAGATGAACAAACAAATGACAGTTTGTGGCT 2456  
2203 Db CATGCTAAATTAATAATCAAGCTATATAACAGATGAACAAACAAATGACAGTTTGTGGCT 2262  
2457 QY CCACAGGAAAACAGGTTCAATAAAGCATCTTCCCAAACAGCTTAGGAGTGTCTGAAGA 2516  
2263 Db CCACAGGAAAACAGGTTCAATAAAGCATCTTCCCAAACAGCTTAGGAGTGTCTGAAGA 2322  
2517 QY TTGACAGGTTGACTTTTCTGCAAGTGAATGATGAATGATGAATGATGAATGATGAATGAT 2576  
2323 Db TTGACAGGTTGACTTTTCTGCAAGTGAATGATGAATGATGAATGATGAATGATGAATGAT 2382  
2577 QY AATCCACCCCTGGACTTTGGAGACACAGCAAGATTTTAGAGTGGAACTCACACCCCAAAA 2636  
2383 Db AATCCACCCCTGGACTTTGGAGACACAGCAAGATTTTAGAGTGGAACTCACACCCCAAAA 2442  
2637 QY ACCATAGCGGAAATGTGACAAAGAAAAGCCCCCATCTCTGATTGTTCCACTGGAAAGC 2696

2443 Db ACCATAGCGGAAATGTGACAAAGAAAAGCCCCATCTCTGATTGTTCCACTGGAAAGC 2502  
2697 QY CAGATGACAAAGAAAAGAAAATCAACAGGAAAAGAAAAGAGAACAGTAGAATGGAGAA 2756  
2503 Db CAGATGACAAAGAAAAGAAAATCAACAGGAAAAGAAAAGAGAACAGTAGAATGGAGAA 2562  
2757 QY AATGCTGAAAATCACAATAGGCGTTACTGAAGTGTACTTGAAGAAAGCTCAGCATTAC 2816  
2563 Db AATGCTGAAAATCACAATAGGCGTTACTGAAGTGTACTTGAAGAAAGCTCAGCATTAC 2622  
2817 QY ACAGATAGTTACTTGGGCTTTTGGCATGGGAGAAAAGAAAAGTATTTTCAAGATCTTCTC 2876  
2623 Db ACAGATAGTTACTTGGGCTTTTGGCATGGGAGAAAAGAAAAGTATTTTCTAGATCTTCTC 2682  
2877 QY GACGAGAGAGTCAATGAGACACAATTTGGCATATGTAATGTAATGTAATGTAATGTAATG 2936  
2683 Db GACGAGAGAGTCAATGAGACACAATTTGGCATATGTAATGTAATGTAATGTAATGTAATG 2742  
2937 QY AGGCAACTAAAAGATACATTTTGCAGATTCCTTCAGATATGTAATGTAATGTAATGTAATG 2996  
2743 Db AGATACAGAGAGATACATTTTGCAGATTCCTTCAGATATGTAATGTAATGTAATGTAATG 2802  
2997 QY AAGTTTGGATTACATCGGGAAGTCCCTGCTCAGATGCTCAGATGCTCAGATGCTCAGATG 3056  
2803 Db AAGTTTGGATTACATCGGGAAGTCCCTGCTCAGATGCTCAGATGCTCAGATGCTCAGATG 2862  
3057 QY GTTATGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAGACGCTCAATTTACAAA 3116  
2863 Db GTTATGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAGACGCTCAATTTACAAA 2922  
3117 QY GTGCGCATTTCTGAGGATATGCAAGTTTGCCTTCTTATTTTATTTATTTCTCATGAGTCA 3176  
2923 Db GTGCGCATTTCTGAGGATATGCAAGTTTGCCTTCTTATTTTATTTATTTCTCATGAGTCA 2982  
3177 QY GTGACGACCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTGT 3236  
2983 Db GTGACGACCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTGT 3042  
3237 QY TTGCTCTGACAGAGAAATCCGAACTGCTACCAAGATTCACAGATGCTGCTTAAAGTTT 3296  
3043 Db TTGCTCTGACAGAGAAATCCGAACTGCTACCAAGATTCACAGATGCTGCTTAAAGTTT 3102  
3297 QY CAGGATTTGACAGGCTCGGAACACATGCTAATTAATTTGCTCAAAATGCTTCTGCTGAT 3356  
3103 Db CAGGATTTGACAGGCTCGGAACACATGCTAATTAATTTGCTCAAAATGCTTCTGCTGAT 3162  
3357 QY ATCAGCGCAGTAAATTAATTTCCACCACTCAGGAATCTTACTATGATCCCAACCTGCCA 3416  
3163 Db ATCAGCGCAGTAAATTAATTTCCACCACTCAGGAATCTTACTATGATCCCAACCTGCCA 3222  
3417 QY CCGGTCTCTAAAGTCTAGTAAACAACTGTAACAACTGTAACAACTGTAACAACTGTAAC 3476  
3223 Db CCGGTCTCTAAAGTCTAGTAAACAACTGTAACAACTGTAACAACTGTAACAACTGTAAC 3282  
3477 QY TATAAGGACAAAACAAATATAGTTTGAATTCATGGGAGAGAAAGAAATCGCTTTTAAA 3536  
3283 Db TATAAGGACAAAACAAATATAGTTTGAATTCATGGGAGAGAGAAATCGCTTTTAAA 3342  
3537 QY ATGATTCGTACCAAGTTTCTCATGTGTTGGCAGTTTGGATGACATAAGAAAACCCCT 3596  
3343 Db ATGATTCGTACCAAGTTTCTCATGTGTTGGCAGTTTGGATGACATAAGAAAACCCCT 3402  
3597 QY AGGAAGTTTGTGCTGAATGACAACTTCCACCACTGTAACAACTGTAACAACTGTAACAA 3656  
3403 Db AGGAAGTTTGTGCTGAATGACAACTTCCACCACTGTAACAACTGTAACAACTGTAACAA 3462  
3657 QY AAGCTGTCTCAGGACTTCTATGAATTCATGTTTCCCACTACCTTCCCAATTTGAATG 3716  
3463 Db AAGCTGTCTCAGGACTTCTATGAATTCATGTTTCCCACTACCTTCCCAATTTGAATG 3522  
3717 QY CCAGAGAGTATCGAAACCGTTTCTCATATGATGCTGAGGAGTATGGAGGCTTAT 3776

Db 3523 CCAAGAGAGTATCGAAACCGTTTCCTTCATATGATGATGCTGAGGAATGGAGGCTTAT 3582  
QY 3777 CGAGACAAATGAAGT 3792  
Db 3583 CGAGACAAATGAAGT 3598  
RESULT 10  
AAD62490  
ID AAD62490 standard; DNA; 3600 BP.  
AC AAD62490;  
XX  
DT 15-JAN-2004 (first entry)  
XX  
DE N-acetylglucosamine-1 (GlcNAc)-phosphotransferase cDNA.  
XX  
KW Mannose glycoprotein; gene therapy; carbohydrate deficient cell;  
KW lysosomal storage disease; N-acetylglucosamine-1-phosphotransferase;  
KW gastrointestinal; gene; ds; lectin resistant cell; deoxymannojirimycin;  
KW kifunensine; glycosylation inhibition.  
XX  
OS Unidentified.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..3600  
FT /\*tag= a  
FT /product= "N-acetylglucosamine-1 (GlcNAc)-  
FT phosphotransferase"  
XX  
PN US2003124652-A1.  
XX  
PD 03-JUL-2003.  
XX  
PF 21-DEC-2001; 2001US-00023889.  
XX  
PR 21-DEC-2001; 2001US-00023889.  
XX  
PA (NOVA-) NOVAZYME PHARM INC.  
XX  
PI Canfield WM;  
XX  
DR WPI; 2003-810984/76.  
DR P-PSDB; ABW01487.  
XX  
XX  
PT Producing a high mannose glycoprotein for treating lysosomal storage  
PT disease, comprises culturing the lectin resistant mammalian cell in the  
PT presence of deoxymannojirimycin and kifunensine.  
XX  
PS Claim 11; Page 8-10; 46pp; English.  
XX  
CC The invention relates to a method for producing a high mannose  
CC glycoprotein. The method comprises: introducing and expressing a  
CC polynucleotide encoding a glycoprotein into a mammalian cell; culturing  
CC the cell in the presence of a lectin to obtain a lectin resistant cell;  
CC isolating the cell; culturing the cell in the presence of  
CC deoxymannojirimycin and kifunensine to inhibit glycosylation of the  
CC glycoprotein; and collecting the glycoprotein. The invention is useful in  
CC gene therapy. The method is useful for producing a high mannose  
CC glycoprotein in a complex carbohydrate deficient cell for treating  
CC lysosomal storage disease. The present sequence is N-acetylglucosamine-1  
CC (GlcNAc)-phosphotransferase cDNA  
XX  
SQ Sequence 3600 BP; 1151 A; 736 C; 798 G; 915 T; 0 U; 0 Other;  
Query Match 61.9%; Score 3465.6; DB 10; Length 3600;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 3477; Conservative 0; Mismatches 19; Indels 0; Gaps 0;  
QY 297 AGCCGAGATCAATACATGTTTGTTCCTATAGACATATTCCTGGAAGTCC 356  
Db 103 AGCCGAGATCAATACATGTTTGTTCCTATAGACATATTCCTGGAAGTCC 162

QY 357 TTTCAAAATCGGCTTTGTCTGCCCATGCCGATTGAGTTGTTTACACCTGGGTGATGCG 416  
Db 163 TTTCAAAATCGGCTTTGTCTGCCCATGCCGATTGAGTTGTTTACACCTGGGTGATGCG 222  
QY 417 ACAGATCTTGAATCTACTGAAGGAACTACAGCAGGCTCAGAGAAACAGATGAGGAGGAGCAG 476  
Db 223 ACAGATCTTGAATCTACTGAAGGAACTACAGCAGGCTCAGAGAAACAGATGAGGAGGAGCAG 282  
QY 477 AAAGCAATGAGAGAAATCCTTGGGAAACACAAACGGAACCTACTAAGAGATGAGAG 536  
Db 283 AAAGCAATGAGAGAAATCCTTGGGAAACACAAACGGAACCTACTAAGAGATGAGAG 342  
QY 537 CAGTTAGAGTGTTCCTAACAACATGCTAATTAAGGCTGCCAATGCTTGTACTGGACCCAGCC 596  
Db 343 CAGTTAGAGTGTTCCTAACAACATGCTAATTAAGGCTGCCAATGCTTGTACTGGACCCAGCC 402  
QY 597 CTGCCAGCAACATACCTGAGGAGCTGCCATCTCTTTATCTCTTTTCTCTTTTCTCTGCC 656  
Db 403 CTGCCAGCAACATACCTGAGGAGCTGCCATCTCTTTATCTCTTTTCTCTTTTCTCTGCC 462  
QY 657 AGTGACATTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCAGTTGTTGT 716  
Db 463 AGTGACATTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCAGTTGTTGT 522  
QY 717 TTTGACAGTACTAAGGATGTTGAAGATGCCCACTCTGACTGCTTAAAGGAAATAGCAGA 776  
Db 523 TTTGACAGTACTAAGGATGTTGAAGATGCCCACTCTGACTGCTTAAAGGAAATAGCAGA 582  
QY 777 CAGACAGTATGAGGAGGCTACTTGACAAACAGATAAAGAGTCCCTGATAGTCTTAATG 836  
Db 583 CAGACAGTATGAGGAGGCTACTTGACAAACAGATAAAGAGTCCCTGATAGTCTTAATG 642  
QY 837 CAAAGTTTGGCTTTCTGAGTGGATTTCACCAACATTCAGAGGAAACAAATCAACTAAAA 896  
Db 643 CAAAGTTTGGCTTTCTGAGTGGATTTCACCAACATTCAGAGGAAACAAATCAACTAAAA 702  
QY 897 ACAAATTCGCAAGAAATCTTTCTCTAAAGTCAAACTGTTGAGTGTATTCAGAGGCC 956  
Db 703 ACAAATTCGCAAGAAATCTTTCTCTAAAGTCAAACTGTTGAGTGTATTCAGAGGCC 762  
QY 957 AGTGAGCGCTTCTAAACCTGAATAACCCCAAGGATTTTCAAGAAATTAAGCAAACT 1016  
Db 763 AGTGAGCGCTTCTAAACCTGAATAACCCCAAGGATTTTCAAGAAATTAAGCAAACT 822  
QY 1017 AAGAAGAACATGACCAATTCATGAAAGAACTGACCATAGTCTGTCATATTTATG 1076  
Db 823 AAGAAGAACATGACCAATTCATGAAAGAACTGACCATAGTCTGTCATATTTATG 882  
QY 1077 GATCTGAGCGCATCAGCCAGTCTAGCAGGATGAGACATCTCTGCCAGTGGTTTGA 1136  
Db 883 GATCTGAGCGCATCAGCCAGTCTAGCAGGATGAGACATCTCTGCCAGTGGTTTGA 942  
QY 1137 GATAACGAAACTGAGGTACTCATTTGCGATCTATCGAGAGGATGACCATGGGTTCG 1196  
Db 943 GATAACGAAACTGAGGTACTCATTTGCGATCTATCGAGAGGATGACCATGGGTTCG 1002  
QY 1197 AATATTTTATGTCACCAACGGGAGATTCATCTGCTGTAACCTTGACATCTCGA 1256  
Db 1003 AATATTTTATGTCACCAACGGGAGATTCATCTGCTGTAACCTTGACATCTCGA 1062  
QY 1257 GTCACAATATGACACACAGGATGTTTTCGAAATTTGAGCCACTGCTACCTTTAGT 1316  
Db 1063 GTCACAATATGACACACAGGATGTTTTCGAAATTTGAGCCACTGCTACCTTTAGT 1122  
QY 1317 TCACCTGCTATTGAAAGTCAATTCATCGATCGAAGGCTGTCACAGAGTTTATTAC 1376  
Db 1123 TCACCTGCTATTGAAAGTCAATTCATCGATCGAAGGCTGTCACAGAGTTTATTAC 1182  
QY 1377 CTAATATGATGATGATGTTTGGAGAGGATGCTGCCAGATGATTTTACAGTCACTCC 1436  
Db 1183 CTAATATGATGATGATGTTTGGAGAGGATGCTGCCAGATGATTTTACAGTCACTCC 1242  
QY 1437 AAAGGCCAGAGGTTTATTGATGATGCTGTGCCAAACTGTGCCGAGGCTGCCAGGT 1496





QY 1437 AAGCCAGAGGTTTATTTGACATGGCTGTGCCAAACTGTGCGGAGGCTGCCAGGT 1496  
DB 1243 AAGGCCAGAGGTTTATTTGACATGGCTGTGCCAAACTGTGCGGAGGCTGCCAGGT 1502  
QY 1497 TCCTGGATTAAAGGATGGCTATTTGTGCAAGGCTTGTAAATAATTCAGGCTCGGATTTGGAT 1556  
DB 1303 TCCTGGATTAAAGGATGGCTATTTGTGCAAGGCTTGTAAATAATTCAGGCTCGGATTTGGAT 1562  
QY 1557 GGTGGGATTTGCTCTGGAAAACAGTGGAGGAGTGCCTATATTTGACAGAGTGGAGTACT 1616  
DB 1363 GGTGGGATTTGCTCTGGAAAACAGTGGAGGAGTGCCTATATTTGACAGAGTGGAGTACT 1422  
QY 1617 GGGAGTATTTGAGTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTTTAC 1676  
DB 1423 GGGAGTATTTGAGTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTTTAC 1482  
QY 1677 TGTAAATCAGGATGTGGAAATTCCTGGCTCGCTGATTAAGTTCTGTGACCAAGCATGCAAT 1736  
DB 1483 TGTAAATCAGGATGTGGAAATTCCTGGCTCGCTGATTAAGTTCTGTGACCAAGCATGCAAT 1542  
QY 1737 GTCTTGCTCTGGGTTTGTGCTGGGACAGTGTGGGCAAGATCAATTTTCATGAATTTGAT 1796  
DB 1543 GTCTTGCTCTGGGTTTGTGCTGGGACAGTGTGGGCAAGATCAATTTTCATGAATTTGAT 1602  
QY 1797 AAAGTGATTCCTTCCCAACACAGTCACTATATTAATTCAAAAGGTGAATGCCCTGCCT 1856  
DB 1603 AAAGTGATTCCTTCCCAACACAGTCACTATATTAATTCAAAAGGTGAATGCCCTGCCT 1662  
QY 1857 TATTTACGTTTGCAGAGTAGCCAAAGAGGAGTTGAAAGTGCCTATAGTGAATTCCTCA 1916  
DB 1663 TATTTACGTTTGCAGAGTAGCCAAAGAGGAGTTGAAAGTGCCTATAGTGAATTCCTCA 1722  
QY 1917 ATAATTCGACATGCTTCTATTTGCCAAAGTGGAACCCATCCACCTCATAAATGCACAGT 1976  
DB 1723 ATAATTCGACATGCTTCTATTTGCCAAAGTGGAACCCATCCACCTCATAAATGCACAGT 1782  
QY 1977 GGAATGAATGCCACCAATATCTTTTAACTCAAGTTTCAAAATACAAACGATGAAGAG 2036  
DB 1783 GGAATGAATGCCACCAATATCTTTTAACTCAAGTTTCAAAATACAAACGATGAAGAG 1842  
QY 2037 TTTCAAAATGCAGATAACAGTGGAGTGGACACAGGAGGAGGCCAAACCTGAATTTCTACG 2096  
DB 1843 TTTCAAAATGCAGATAACAGTGGAGTGGACACAGGAGGAGGCCAAACCTGAATTTCTACG 1902  
QY 2097 GCCCAGAGGTTTACGAAAATTTAGTTAGTCCCAATACATCTTCTCCAGAGGCGGAAATC 2156  
DB 1903 GCCCAGAGGTTTACGAAAATTTAGTTAGTCCCAATACATCTTCTCCAGAGGCGGAAATC 1962  
QY 2157 CTTTTTGAGGATATTTCCCAAGAAAACGCTTCCCGAAGTTTAAAGACATGATGTTAAC 2216  
DB 1963 CTTTTTGAGGATATTTCCCAAGAAAACGCTTCCCGAAGTTTAAAGACATGATGTTAAC 2022  
QY 2217 TCAACAGGAGAGCCCGAGGAGGTGAAAATTTCCCTGGTAAATATTTCACTCCTTCCA 2276  
DB 2023 TCAACAGGAGAGCCCGAGGAGGTGAAAATTTCCCTGGTAAATATTTCACTCCTTCCA 2082  
QY 2277 AAAGACCCAGTTGAGTCTCMATACCTTGGATTTGCACTGGAACATGAGACATCACT 2336  
DB 2083 AAAGACCCAGTTGAGTCTCAATACCTTGGATTTGCACTGGAACATGAGACATCACT 2142  
QY 2337 TTGAAAGGATACAAATTTGTCCAGTCAAGCTTGTGAGATCAATTTCTGATGAATCACTCACAG 2396  
DB 2143 TTGAAAGGATACAAATTTGTCCAGTCAAGCTTGTGAGATCAATTTCTGATGAATCACTCACAG 2202  
QY 2397 CATGCTAAAATAAAAAATCAAGCTATATAACAGATGAACAAATGACAGTTTGGTGGCT 2456  
DB 2203 CATGCTAAAATAAAAAATCAAGCTATATAACAGATGAACAAATGACAGTTTGGTGGCT 2262  
QY 2457 CCACAGAAAAACAGGTTTCATAAAGCATCTTGGCAAAACAGCTTAGGAGTGTCTGAAAGA 2516  
DB 2263 CCACAGAAAAACAGGTTTCATAAAGCATCTTGGCAAAACAGCTTAGGAGTGTCTGAAAGA 2322  
QY 2517 TTGCAGAGGTTGACTTTTCTCTGCGAGTGAAGTGTAAAAAGTGAATGGTCAATGACCGAGGTCAG 2576

DB 2323 TTGCGAGAGTTGACTTTTCTCTGAGTGAGTGTAAAGTGAATGTCATGACAGAGGTCAG 2382  
QY 2577 AATCCACCCCTGGACTTTGGAGACCAACAGAAAGATTTAGAGTGGAACTCACACCCAAAAA 2636  
DB 2383 AATCCACCCCTGGACTTTGGAGACCAACAGAAAGATTTAGAGTGGAACTCACACCCAAAAA 2442  
QY 2637 ACCATAGCGGAAATGTCACAAAGAAAGAGCCCATCTCTGATTTGTTTCCACTGGAAGC 2696  
DB 2443 ACCATAGCGGAAATGTCACAAAGAAAGAGCCCATCTCTGATTTGTTTCCACTGGAAGC 2502  
QY 2697 CAGATGACAAAAGAAAGAAATCACAGGGAAGAAAGAAAGAGACAGTAGAATGGAGAA 2756  
DB 2503 CAGATGACAAAAGAAAGAAATCACAGGGAAGAAAGAAAGAGACAGTAGAATGGAGAA 2562  
QY 2757 AATGCTGAATATCATATAGGCGTTACTGAAGTGTACTTGGGAAGAAAGCTGCGCATTTAC 2816  
DB 2563 AATGCTGAATATCATATAGGCGTTACTGAAGTGTACTTGGGAAGAAAGCTGCGCATTTAC 2622  
QY 2817 ACAGATAGTTACTTGGGCTTTTGGCATGGGAGAAAAAGATTTTCCAAAGATCTTCTC 2876  
DB 2623 ACAGATAGTTACTTGGGCTTTTGGCATGGGAGAAAAAGATTTTCCCTAGATCTTCTC 2682  
QY 2877 GACGAAGAGAGTONTTGAAGACACAAATTTGGCATACTTCTGATAGCAAAAAATCTGGG 2936  
DB 2683 GACGAAGAGAGTONTTGAAGACACAAATTTGGCATACTTCTGATAGCAAAAAATCTGGG 2742  
QY 2937 AGGCAACTTAAAGATACATTTTGCAGATTCCTCTCAGATGTAAATAAAATTTCTAAATAGC 2996  
DB 2743 AGATCAAGAGAGATACATTTTGCAGATTCCTCTCAGATGTAAATAAAATTTCTAAATAGC 2802  
QY 2997 AAGTTTGGATTCACATCCGGAAGTCCCTGCTCACATGCTCACATGATTTGACCGGAT 3056  
DB 2803 AAGTTTGGATTCACATCCGGAAGTCCCTGCTCACATGCTCACATGATTTGACCGGAT 2862  
QY 3057 GTTATGCAAGACTGCAAGTATGTTCCCTGAAGATTTGCAAGAGTGCATTTTCAAAA 3116  
DB 2863 GTTATGCAAGACTGCAAGTATGTTCCCTGAAGATTTGCAAGAGTGCATTTTCAAAA 2922  
QY 3117 GTGCGCCATTTCTGAGGATATGAGTTTGCCTTCTCTTATTTTATTTATCTCATGAGTGA 3176  
DB 2923 GTGCGCCATTTCTGAGGATATGAGTTTGCCTTCTCTTATTTTATTTATCTCATGAGTGA 2982  
QY 3177 GTGCGCCACTGATATATCTCAAGTCTTTGATGAAGTTGATACAGATCATCTGGTGTG 3236  
DB 2983 GTGCGCCACTGATATATCTCAAGTCTTTGATGAAGTTGATACAGATCATCTGGTGTG 3042  
QY 3237 TTGCTGACAGAGAAATCCGGAACACTGGCTACCAAGATTTACGAACTGCCGTTAAGTTTG 3296  
DB 3043 TTGCTGACAGAGAAATCCGGAACACTGGCTACCAAGATTTACGAACTGCCGTTAAGTTTG 3102  
QY 3297 CAGGATTTGACAGGTTCTGGAAACATGCTATAATAAATTTGCTCAAAAATGCTTCTGCTGAT 3356  
DB 3103 CAGGATTTGACAGGTTCTGGAAACATGCTATAATAAATTTGCTCAAAAATGCTTCTGCTGAT 3162  
QY 3357 ATCAGCGAGCTAAATAATTTTCCCAACACTCAGGAATCTTACTATGATCCCAACCTGCA 3416  
DB 3163 ATCAGCGAGCTAAATAATTTTCCCAACACTCAGGAATCTTACTATGATCCCAACCTGCA 3222  
QY 3417 CCGGTCACTAAAAGTCTAGTAAACAAACTGTAAACAGTAACTGACAAAATCCACAAAGCA 3476  
DB 3223 CCGGTCACTAAAAGTCTAGTAAACAAACTGTAAACAGTAACTGACAAAATCCACAAAGCA 3282  
QY 3477 TATAGGACAAAACAAATATAGGTTTGAATTCATGGGAGAGAAATTCGCTTTTAAA 3536  
DB 3283 TATAGGACAAAACAAATATAGGTTTGAATTCATGGGAGAGAAATTCGCTTTTAAA 3342  
QY 3537 ATGATTCCTACCAAGTTTCTCATGTGGTGGCCAGTTGGATGATCATAGAAAACCCCT 3596  
DB 3343 ATGATTCCTACCAAGTTTCTCATGTGGTGGCCAGTTGGATGATCATAGAAAACCCCT 3402  
QY 3597 AGGAAGTTTGTTCCTGGAATGACAAACATTTGACCAAAATCATTAAGATGCTCAGACAGT 3656



Db 3403 AGGAAGTTTGTTCCTGTAAGTACCAACATTTGACCAACATATTAAGATGCTACAGACAGTG 3461  
 Qy 3657 AAGCTGTCTCAGGAGACTTCTATGAATCCATGTTCCTCCATACCTTCCCAATTGAACTG 3716  
 Db 3463 AAGCTGTCTCAGGAGACTTCTATGAATCCATGTTCCTCCATACCTTCCCAATTGAACTG 3522  
 Qy 3717 CCAAGACAGTATCGAAACCGTTTCCTTCATATGCAATGAGCTGCAGGAATCGAGGCGCTTAT 3776  
 Db 3523 CCAAGACAGTATCGAAACCGTTTCCTTCATATGCAATGAGCTGCAGGAATCGAGGCGCTTAT 3582  
 Qy 3777 CGAGACAAATTTGAAGT 3792  
 Db 3583 CGAGACAAATTTGAAGT 3598  
  
 RESULT 12  
 ACC81007  
 ID ACC81007 standard; cDNA; 5229 BP.  
 XX  
 AC ACC81007;  
 XX  
 DT 01-AUG-2003 (first entry)  
 XX  
 Murine GlcNAc-phosphotransferase alpha/beta-subunit cDNA.  
 XX  
 Mouse; N-acetylglucosamine-1-phosphotransferase; nephrotropic;  
 KW GlcNAc-phosphotransferase; phosphodiester alpha-GlcNAcase; Gene; ss;  
 KW N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase;  
 KW enzyme replacement therapy; phosphorylated lysosomal hydrolase;  
 KW lysosomal storage disease; enzyme; beta-subunit; alpha-subunit.  
 XX  
 Mus musculus.  
 OS  
 XX  
 US6537785-B1.  
 PN  
 XX  
 25-MAR-2003.  
 PD  
 XX  
 10-AUG-2000; 2000US-00636077.  
 PF  
 XX  
 14-SEP-1999; 99US-0153831P.  
 PR  
 XX  
 (GENZ-) GENZYME GLYCOBIOLOGY RES INST INC.  
 PA  
 XX  
 Canfield WM;  
 PI  
 XX  
 WPI; 2001-290356/30.  
 DR  
 P-PSDB; ABR61380, ABR61384.  
 XX  
 Novel N-acetylglucosamine-1-phosphotransferase and N-acetylglucosamine-1-  
 PT phosphodiester alpha-N-Acetylglucosaminidase, useful for producing  
 PT phosphorylated lysosomal hydrolase for treating lysosomal storage  
 PT diseases.  
 XX  
 PS Disclosure; Page 47-49; 62pp; English.  
 XX  
 CC The invention relates to a novel isolated human N-acetylglucosamine-1-  
 CC phosphotransferase (GlcNAc-phosphotransferase) (I) and phosphodiester  
 CC alpha-GlcNAcase (N-acetylglucosamine-1-phosphodiester alpha-N-  
 CC Acetylglucosaminidase) (II). The protein of the invention has  
 CC nephrotropic activity, and may be useful in enzyme replacement therapy. A  
 CC protein of the invention (I), (II) is useful for preparing a  
 CC phosphorylated lysosomal hydrolase. The phosphorylated hydrolase  
 CC comprising a terminal mannose-6-phosphate, is useful for treating a  
 CC patient suffering from a lysosomal storage disease. The present sequence  
 CC is used in the exemplification of the invention  
 XX  
 SQ Sequence 5229 BP; 1485 A; 1242 C; 1199 G; 1302 T; 0 U; 1 Other;  
  
 Query Match 50.4%; Score 2820.8; DB 4; Length 5229;  
 Best Local Similarity 79.1%; Pred. No. 0;  
 Matches 3667; Conservative 1; Mismatches 833; Indels 137; Gaps 21;  
 QY 150 GGCGGTGAAGGGGTGATGCTGTTCAGAGCTCTCGACAGACAAACCTATACCTCCCTGTGCC 209

Db	1	GGCGGTGAAGGGGTGATGCTGTTCAAGCTCCTGCAGAGACAGACCTATACCTGCCTATCC	60
QY	210	CACAGTATGGCTCTACGTGTGCTTCTGGGGGTGGTGTGCACCATCTCTCCSCCTTC	269
Db	61	CACAGTATGGCTCTACGTCTGCTTCGTGGGGGTGGTGTGCACCATCTCTCGGCTTTC	120
QY	270	CAGTTCGGAGAGTGGTTCGAAATGGAGCCGAGATCAATACCATGTTTGTGTTGATTC	329
Db	121	CAGTTCGGAGAGTGGTTCGAAATGGAGCCGAGATCAGTACCATGTTTGTGTTGATTC	180
QY	330	TATAGAGACAATATGCTGGAAAGTCCTTTCAGAAATCGGCTTTCCTGCCATCGCGATT	389
Db	181	TACAGAGACAATATGCTGGAAATCCTTTTCAGAAATCGGCTTTCCTGCCATCGCAATC	240
QY	390	GACGTGTGTTTACACTCGGTGAAATGGACAGATCTTGAACACTCTGAAGGAACTACAGAG	449
Db	241	GACGTGTGTTTACACTCGGTGAAATGGACACTGACCTTGAACTGCTAAAGAGACTACAGAG	300
QY	450	GTCAGAGACAATGAG	509
Db	301	GTCAGAGAGACATGAG	360
QY	510	ACGGAACTCTAATAAGAGAGTGAAGACAGTAGAGTGTGCTAAACACACTGCATTAAAG	569
Db	361	ACCGAACTCTAATAAGAGAGTGAAGACAGTAGAGTGTGCTAAACACACTGCATTAAAG	420
QY	570	GTCGCAATGCTTGACTGGACCCAGCCCTGCCAGCCAAATCACCTCGAAGGACGTGCCA	629
Db	421	GTCGCCATGCTTGCTTGACCCCGGCCCTGCCAGCCACCATCACCTGAAGGATCTGCCA	480
QY	630	TCTCTTTATCCTCTTTTCACTTGCCAGTGACATTTTCAATGTCGAAACCAAAAC	689
Db	481	ACCTTTTACCCTATTTTCCAGCGGTCGAGCGACATGTTCAATGTCGAAACCAAAAC	540
QY	690	CCTTCTACCAATGTCAGTGTGTTGTTTGAAGTACTAAGGATGTTGAAGATGCCAC	749
Db	541	CGCTCTACCAATGTCGCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG	600
QY	750	TCTGACGTCTTAAGGAAATAGCAGACAGACAGTATGAGGGGGTACTTGCACACAGAT	809
Db	601	GCTGACCGTTTAAAGGAGCCAGCAACACAGATGTTTGAGAGCCTTCTTGACACAGAC	660
QY	810	AAAGAAATGCTGAAATGCTGCTAAATGCAAGATTTGGCTTTCCTGAGTGGATTTCCACCA	869
Db	661	AAAGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	720
QY	870	ACATTCAGGAGAAACAATCACTAAACAAACAAATTCGCAAGAAATCTTTCTCTCTCTCT	926
Db	721	ACCTTTCAGGAGAGCAGTCACTCACTGAAGCAAAAGCTGCCAAGAAAGCTTTCCTCTAAA	780
QY	927	GTCAACTGTTGAGTGTGTTTACAGAGCCAGTGTAGCGCTTCTAAACCTGAATTAACCC	986
Db	781	ATPAAAGCTGTTGCGCTGCTACTCGGAGGCCAGTGTGCTCTTCTGAAATGAAATATCC	840
QY	987	AAGAAATTTCAAGAAATTTGAATAGCAAACTAAGAGAAACATCACCATTTGATGGAAAGAA	1046
Db	841	AAGGGTTTCCAGAGCTGAAACAGACAGACCAAGAGAAACATCACCATCGATGGGAAGAA	900
QY	1047	CTGACCAATAGCTCTGATATTTATATATATATATATATATATATATATATATATATATAT	1106
Db	901	CTGACCAATAGCTCTGATATCTGCTGTGGAGCCTGAGTGCCATCAGCCAGTCCAGCAG	960
QY	1107	GATGAGACATCTCTGCCAGTCTGTTTGAAGATAAGAGAACTGAGGTACTGATCTGCGA	1166
Db	961	GATGAGACGCTCTGCCAGCCCTTGAAGATAAGAGAGCTGAGGTACTCGCTGCGA	1020
QY	1167	TCTATCAGAGGATGACCATGGTTCGGAATATTTTCATTTGTCACCAAGGGCAGATT	1226
Db	1021	TCTATCAGAGACACGGCCATGGGTACGGAATATTTTCATTTGTCACCAAGGGCAGATT	1080
QY	1227	CCATCTCTGGCTGAACCTTGAACAATCTCGAGTGACAATAGTAACACACAGGATGTTTT	1286

Query Match 50.4%; Score 2820.8; DB 4; Length 5229;  
Best Local Similarity 79.1%; Pred. No. 0;  
Matches 3667; Conservative 1; Mismatches 833; Indels 137; Gaps 21;  
QY 150 GCGGTGAAGGGGTGATCTGTTCAAGCTCTCGACAGACAAACTATATCTGCTGTCC 209

Db 1081 CCATCCTGGCTGAACCTTGACAAACCTCGAGTGACCATAGTGACCCACAGGACATTTTC 1140  
Qy 1287 CGAAATTTGAGCCACTTGGCTCACTCTTGTAGTTACCTCTCTCTATTGAAAGTCACTTATCGC 1346  
Db 1141 CAAAATCTGAGCCACTTGGCTTACTTTTCAGTTCCCTCTCTATTGAAAGTCACTTATCCCGC 1200  
Qy 1347 ATCGAAGGGCTGCTCCAGAGTTTATTTACTCTAATCATGATGTCATGTTTGGGAAGGAT 1406  
Db 1201 ATCGAAGGGCTGCTCCAGAGTTTATTTATCTAATGACGATGTCATGTTTGGTAAGGAC 1260  
Qy 1407 GTCTGGCCAGATGATTTTACAGTCACTTCCAAAGGCCAGAGTTTATTTGATGAGCCT 1466  
Db 1261 GTCTGGCCAGGATTTTACAGCACTTCCAAAGGTCAAAAGGTTTATTTGACATGGCCT 1320  
Qy 1467 GTGCCAACTGTGCGAGGGTGGCCAGGTTCTCTGGATTAAAGGATGGCTATTGTGCAAG 1526  
Db 1321 GTGCCAACTGTGAGAGGGTGGCCGGGCTCTCTGGATAAAGGACGGCTATTGTGTAAG 1380  
Qy 1527 GCTTGTAAATATTCAGCTCGGATTTGGGATGGTGGGATTTGCTCTGGAACAGTGGAGGG 1586  
Db 1381 GCTTGTAAATATTCAGCTCGGATTTGGGATGGTGGGATTTGCTCTGGAATACTGCAAGG 1440  
Qy 1587 AGTGGCTATATTGAGAGGGTGGAGGTACTGGGAGTATTGGAGTTGGACACCCCTGGCAG 1646  
Db 1441 AACCCTGTTTGTGCAAGAGGTGGGGTACCGGGATATTGGAGCTGGACAGCACTGGCAG 1500  
Qy 1647 TTTGGTGAGGAATAAACAAGTGTCTTACTGTATCAGGATGTGCGAATTTCTGGCTC 1706  
Db 1501 TTTGGTGAGGAATAAACAAGTGTCTTACTGTATCAGGATGTGCGAATTTCTGGCTG 1560  
Qy 1707 GCTGATAGTCTGTGACCAAGATGCAATGCTCTTGTCTGTGGTTTGTATGCTGGGAC 1766  
Db 1561 GCTGACAGTCTGTGACCAAGCTGTAACTCTTATCTCTGGGTTTGTATGCTGGTAC 1620  
Qy 1767 TGTGGGCAAGATCAATTTTCAATGATGATTAATGATGATCTTCTCCAAACCAAGCTCAC 1826  
Db 1621 TGTGACAGGATCAATTTTCAATGATGATTAATGATGATCTTCTCCAAACCAAGCTCAC 1680  
Qy 1827 TATATTTTCCAAAGGTGAATGCTCTTATTTACAGTTTTCAGAAAGTGGACCAAGA 1886  
Db 1681 TATGTTGTCCTCCAAAGGTGAATGCTCTTATTTACAGTTTTCAGAAACATAGCCAGAAA 1740  
Qy 1887 GGAGTTGAAGTGTCTATAGTACCAATCCATATTTGACATGCTCTTATTTGCCAAG 1946  
Db 1741 AGAATTGAAGGAGCTTACAGGCAACCCCATCTATCCGCCACCGCTCCATTTGCAACAAG 1800  
Qy 1947 TGGAAACCATCCACCTCATATGACAGTGGAAATGAATGCGACCAACAATACATTTTAAAT 2006  
Db 1801 TGGAAACCATCCACCTCATATGATGATGCGGGGGATGAACGCCACCATATTTTAAAC 1860  
Qy 2007 CTCAGTTTCAAAATACAAAGATGAAGTTTCAAAATGCGATGATTAAGTGGAGTGGAC 2066  
Db 1861 CTCACCTTCTCAAAAGCCCAAGAGTTCAGATCCAGATAGCAGTAGAGTGGAC 1920  
Qy 2067 ACAAGGAGGACCAAACTGAAATTTCTACGCGCCAGAGGGTTACGAAATTTTATGTTAGT 2126  
Db 1921 ACGAGGAGGCGCCCAAACTGAAATTTCTACACCCAGAGGCTATGAAGTTTGGTTAGC 1980  
Qy 2127 CCCATAACACTTTTCCAGAGCGGAAATCTCTTTTGGAGTATTTCCCAAGAAAACGC 2186  
Db 1981 CCAGTGACACCTTCTCTCAGCTGACGCTCTCTTTTGAAGTGTCCCAAGAAAACGC 2040  
Qy 2187 TTCCCGAGTTTGAAGACATGATGTTAACTCAACAGGAGAGCCAGGAGAGGTGAAA 2246  
Db 2041 TTCCCAAGATCAGAGACATGATGTTAAATGCAACAGGAGATTTCCAGAGAGGTGAAA 2100  
Qy 2247 ATTCCTCTGGTAAATATTCTACTCTCTCCAAAGACGCCAGTTGAGTCTCAATACCTTG 2306  
Db 2101 ATCCCCCGGGTAAATATTCTACTCTCTCCAAAGAGGCCAGGTTGAGCTGAGCAACTTG 2160  
Qy 2307 GATTTGCACTGACATGGAGACATCACTTTGAAAGGATCAATTTGTCCTAGTCAAGC 2366  
Db 2161 GATTTGCACTAGAACCGTGGAGACATCACTCTGAAAGGATTAATCTTGTCCAGTCAAGC 2220

Qy 2367 TTGCTGAGATCATTTCTGATGAATCTCACAGCATGCTAAAAATAAAAATCAAGCTATATA 2426  
Db 2221 CTGCTTAAGTCTTTCTCTGGGAAATCTACTAGATATAAATAAACCTCAAGCT--AGG 2277  
Qy 2427 ACAGATGAAAACAAATGACAGTTTGGTGGCTCCACAGGAAAAACAGGTTTCATAAAGCATC 2486  
Db 2278 ACCGATGAAAACAAAGGCAACCTGGAGTCCACAGGAAAAACCTTCTCTCAGACG--- 2333  
Qy 2487 TTGCCAAAACAGCTTAGGAGTGTCTGAAGATTGACAGAGTTGACTTTCTCTGCAAGTGA 2546  
Db 2334 --TCCACATGGCTTTGCTGGTGAACACAGATCAGAGATGCACTGCCCCAGCAGACACA 2391  
Qy 2547 GTAAAAGTGAATGTCTATGACACAGGCTCAGAAATCCACCCCTGGACTTTGGAGACCACAGCA 2606  
Db 2392 GTGACCGTGAAGGCGGTGACCAAGCTTTGAATCCACCCCGGTGTTGGAGACCAATGCA 2451  
Qy 2607 AGATTTAGAGTGGAAACTCACAACCAAAAACCATAGCGGGAATGTGACAAAAGAAAAG 2666  
Db 2452 AGATT-----GGCCCAAGCTACACTAGCGGTGACTGTCTCCAAAGAGAAC 2496  
Qy 2667 CCCCCATCTCTGATTTGTTCCACTTGAAAAGCCAGATGACAAAAGAAAAGAAAATCACAGG 2726  
Db 2497 CTTTCCACCGCTGATGTTCCCCAGAAAGCCACTTTGCCAAAAGAAAGAGAGAGTGAACAG 2556  
Qy 2727 AAAGAAAAGAGAACAGTAGAATGGAGGAAAATGCTGAAAATCACAATAGGCGTTACTGAA 2786  
Db 2557 GCAGAA-----GGCAATGCTGTACCTGTAAGAG 2586  
Qy 2787 GTGTTACTTGAAGAAAAGCTGCAGCATTAACAGATAGTTACTTTGGGCTTTTGGCATGG 2846  
Db 2587 TTAGTGCCTGGCAGAGCGTTGCAGC-----AGAATTATCCAGGCTTTTGGCCTGG 2637  
Qy 2847 GAGAAAAAAAGTATTTCCAGATCTTCTCGACGAGAGAGTCAATTGACAGACAAATTG 2906  
Db 2638 GAGAAAAAAAGTATTTCCAGACCTTCTGATGAGGAGAGTCAATTGAAAGACCCAGTTG 2697  
Qy 2907 GCATATCTCACTGATGACAAAATACTCGGAGGCACTTAAAGATACATTTTGCAGATTCC 2966  
Db 2698 GCGTACTTTACAGACCGCAACATACCGGAGGCACTTAAAGATACATTTTGCAGACTCC 2757  
Qy 2967 CTCAGATATGTAATAAATAATTTCTAAATAGCAAGTTTGGATTCAATCCGCGAAAGTCCCT 3026  
Db 2758 CTCAGATATGTAATAAATAATTTCTCAACAGCAAGTTTGGATTCAATCCAGGAAAGTCCCT 2817  
Qy 3027 GCTCATGCTCATGATGACCGGATTTGTTAGCAGAGTCTGAGAGTCAAGATATGTTCCCT 3086  
Db 2818 GCACACATGCCCGACATGATGACAGGATCGTTATGCAAGAACTCCAAGATATGTTCCCT 2877  
Qy 3087 GAAGAATTTGACAGACGTCATTTCAAAAGTGCGCCATTTCTGAGGATATGACGTTTGCCT 3146  
Db 2878 GAAGAATTTGACAGACGTCATTTCAAGGTGCGTCACTCTGAGGACATGACGTTTGCCT 2937  
Qy 3147 TTTCTCTATTTTATATCTCATGAGTGCAGTGCAGCCTGAATATATCTCAAGTCTTT 3206  
Db 2938 TTTCTCTACTTTTATTAATCTCATGAGTGCAGTTCAGCCCCCTCAATATTTTCCCAAGTCTTT 2997  
Qy 3207 GATGAGTTGTATACAGATCAATCTGCTGTCTGTCTGACAGAGAAATCCGAAACACTGGCT 3266  
Db 2998 CATGAGTGTAGACAGACCAATCTGCTGTCTGTCTGATAGGAAATCCGAACTGTCGCC 3057  
Qy 3267 ACCAGAAATTCAGAACTGCGTTAAAGTTGACAGGATTTGACAGGTTCTGSAACACATGCTA 3326  
Db 3058 ACCAGAAATTCAGAACTGCTTAAAGCTTGAAGATTTGACAGGTTTGAAGACACATGTTA 3117  
Qy 3327 ATAAATGCTCAAAATGCTTCTCTGATATCAGCAGCTTAAATAATATTTCCACCACT 3386  
Db 3118 ATAAATGCTCAAAATGCTTCTCTGATATATCTCACTCACTCAACACATCCACCGACT 3177  
Qy 3387 CAGGAATCTCTATGATCCCAACCTGCCACCGGTCACTAAAGTCTAGTAACAACTGT 3446  
Db 3178 CAGGAAGCATACTACGACCCCAACCTGCTCGGTCACTAAGAGTCTTGTCAACCACTGT 3237

3447 AAACAGTAACAGTACAAATCCAAAGCATATAAGGACAAACAAATATAGGTTTGAA 3506  
3238 AAGCCAGTAACAGTACAAAGTACCAAGAGCTATATAAGACAAACAAATACAGGTTTGAA 3297  
3507 ATCATGGGAGAGAGAAATCGCTTTTAAATGATTCGTAACAGGTTTCTCATGTTGTT 3566  
3298 ATCATGGGAGAGAGAAATCGCTTTTAAAGATGATACGAACCAATGTTTCTCATGTTGTT 3357  
3567 GGCAGTTGGATGACATAAGAAACCCCTAGGAAGTTTGTGCTGCTGAATGACAAACATT 3626  
3358 GGTCAAGTTGGATGACATACAGAAACCCAGGAGTTGTTGTCTGAATGACAAACATT 3417  
3627 GACCAACATCAATGAAGTCTCAGACAGTGAAGGTTGTTCTCAGGAGCTTCTATGAATCC 3686  
3418 GACCAACATCAATGAAGTCTCAGACAGTGAAGGTTGTTCTCAGGAGCTTCTATGAATCC 3477  
3687 ATGTTCCCATACCTTCCCAATTTGAACCTGCAAGAGAGTATCGAAACCGTTTCCCTTCAT 3746  
3478 ATGTTTCCCATACCTTCCCAATTTGAACCTGCAAGAGAGTATCGAAACCGTTTCTGCAC 3537  
3747 ATGCATGAGCTCGAGGAATGAGGAGGCTTATCGAGACAAATGGAAGTTTGGACCAATGTT 3806  
3538 ATGCATGAGCTCGAGGAATGAGGAGGCTTATCGAGACAAATGGAAGTTTGGACCAATGTT 3597  
3807 GTACTAGCAACATGATGATGTTTACTATATCTCATTTTCTGCTGAGCAGTTAATGCA 3866  
3598 GTACTAGCAACATGATGATGTTTACTATATCTCATTTTCTGCTGAGCAATATGCT 3657  
3867 CTTAAGCGGAAGATATTTCCAGAGGAGGATACACAAAGAGCTAGTCCCAATCGAATC 3926  
3658 CTGAAGCGGAAGATATTTCCAGAGGAGGATACACAAAGAGCTAGTCCAGCGAATC 3717  
3927 AGAGTATAGAGATCTTCAATTTGAAACCATCTACCTCAGACATTTACTGAGCAATTTAAA 3986  
3718 AGGGGTGAAGATCTTCAATTTGAAAGTCACCTAGCTTAGCATCT-GTGAACATCTCCCT 3776  
3987 ACTCAGCTTCAGAGATCTCTTTGTGATGTCATGCTTA-GCAGTTTGCSCCGAAGAGG 4045  
3777 CTTGACACCAAGCGGAGTCTCTGTGATGTCGACAGGAGCGCTCGTGGGAGAGG 3836  
4046 AAAATATCAGTACCATGCTGTTTGTGGCATGATATAGCCACTGACTAGGAATATT 4105  
3837 GACATCGTGCAGACCGGGTCTCTCTGCAATGGGAAGAGAGCCACTGACCTGGAATATT 3896  
4106 TAACCAACCCACTGAAACCTGTGTGTCGAGCAGCTCTGAAGTATTTTACTTTTAAAGA 4165  
3897 ----CAGCACATAGAACCTGTGCAATAGC-----TTGTACAGCTTGTACTTTTAAAGG 3948  
4166 ATTTGCTCATGACCTGTCTATCTTTTATAAAGAGGCTCACTGACAAAGAGACAGCTGTT 4225  
3949 ATTTGCGGAGGACCTGTGCGCTTGTGACAAACCTCCCTGACA-----AGCTGCT 4000  
4226 AATTTCCACAGCAATCATTTGAGACACTAACTTTATAGGAGAGGCTATGCCAGCTGGA 4285  
4001 GGTCTTCTTCCCGGTTACTGAGACT-----GAGAAACCAAGTCCATCTTGAAA 4049  
4286 GTGATTGCTAAGAGCTCAGTCTTGTGATTCCTCAAGGCTTTTGTAAAGTTTGTCACTT 4345  
4050 GCAAGTGGGAGGAGGCGCCAGTCTTGTGATTCCTCAAGGCTTTCCAGCAATTTCTGGCTT 4109  
4346 TT-----TTTTTTTCAATTTCCCAATTTT-----AAGTAGTTTACTAAGTTAACT 4388  
4110 GTCTGCTCTCTTGTGATCCATTTCCCAATTTT-----TATAAACAATAAGTGGCTAAAGTT 4169  
4389 AGTTATTTCTGCTCTGATGATACGAATGCGATGCTTAACCTATTTTATATAGATTT 4448  
4170 AGTCATTTCTCTCTCAAAATAACAAATCAGGATGTCAAAC--ATTGTATAGATCTT 4227  
4449 ATTTAAATAAGCAGCAATATCACCTCTTATTTGACAATACCTAAATATAGATTGTTTATTA 4508  
4228 ATTTAAATAATAGACGATTTACTTCT--TTAGCCTATCAATTTATTTGATTTTATTA 4285  
4509 ATATTAGACTGTAAATGGTCTTAAACACT--AACTACTGAAGAGCTCAATGATTGA 4565

4286 ACAGTCAAG-----TGGTCTTGAACCGCTAACACTACTGAAGAGCTCGA-GATTGA 4336  
4566 CATCTGAATGCTTTGTAATATTGACTTCAGCCCTTGAAGTCTGATGATTTACAGTGC 4625  
4337 CGTTGAAGTCTTTGAGCTTGTAACTCATTTCCCAAGATACTGTGACCTCGTGTGC 4396  
4626 AGGTCTAATTTCAACAGGCTAGAGT-----TAGTACTACTTACCAGATGTAAATTATGT 4678  
4397 GGGCTGATTCGGAAGGCTAGTGTACGTAGCAGTCTCTCACCGGATGTAAATTATGT 4456  
4679 TTTGGAAATGTACATATTCAAACAGAGTCCCTCATTTTGAATAATGAGTAGTCTGATGG 4738  
4457 CGTGGAAATGTACATA--CAGACAAAGTGCCTCACTTCAGAAATGAGTAGTCTGATGG 4514  
4739 CACTGGCACATTACAGTG 4756  
4515 CACCAGCAGTGTGGTG 4532

RESULT 13  
ADD27816  
ID ADD27816 standard; DNA; 5229 BP.  
XX  
AC ADD27816;  
XX  
DT 15-JAN-2004 (first entry)  
XX  
DE GlcNAc-phosphotransferase associated DNA #1.  
XX  
KW mouse; protein phosphorylation; soluble GlcNAc-phosphotransferase;  
XX  
KW UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease; ds.  
XX  
OS Mus musculus.  
XX  
PN US2003119088-A1.  
XX  
PD 26-JUN-2003.  
XX  
PF 21-DEC-2001; 2001US-00023888.  
XX  
PR 21-DEC-2001; 2001US-00023888.  
XX  
PA (NOVA-) NOVAZYME PHARM INC.  
XX  
FI Canfield W, Kudo M;  
XX  
DR WPI; 2003-801323/75.  
XX  
PT Phosphorylating a protein for treating a patient suffering from a  
PT lysosomal storage disease e.g. Fabry's disease by contacting the protein  
PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated  
PT protein.  
XX  
PS Disclosure; SEQ ID NO 8; 55pp; English.  
XX  
CC The invention relates to a method of phosphorylating a protein comprising  
CC contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-  
CC acetylglucosamine) and producing a phosphorylated protein. The method is  
CC useful for treating a patient suffering from a lysosomal storage disease  
CC e.g. Fabry's disease. The present sequence represents a GlcNAc-  
CC phosphotransferase associated DNA.  
XX  
SQ Sequence 5229 BP; 1485 A; 1242 C; 1199 G; 1302 T; 0 U; 1 Other;

Query Match 50.4%; Score 2820.8; DB 10; Length 5229;  
Best Local Similarity 79.1%; Pred. No. 0;  
Matches 3667; Conservative 1; Mismatches 833; Indels 137; Gaps 21;  
Qy 150 GCGGGTGAAGGGGTGATGCTTCAAGCTCCTCGAGACAAACCTATACCTGCTGCTCC 209  
Db 1 GCGGGTGAAGGGGTGATGCTTCAAGCTCCTCGAGACAGACACCTATACCTGCTATCC 60



Db 2221 CTGCTAAGGTCTTCTCGGGGAATTCACGTAGATATAAAATAAAACCTCAAGCT---AGG 2277  
Qy 2427 ACAGATGAACAATGACAGATTTGGTGGCTCCACAGGAAACACAGGTTTCATAAAGCATC 2486  
Db 2278 ACCGATGAACAACAAAGGCAACCTGGAGTCCACAGGAAACCTTCTCACAGACG--- 2333  
Qy 2487 TTGCCAAACAGCTTAGGAGTGTCTGAAAGATTGACAGGTTGACCTTTTCTCGCAGTGAGT 2546  
Db 2334 --TCCACATGGCTTTGCTGGTGAACACACAGATCAGAGAGATGACTGCCCCACGACAGACA 2391  
Qy 2547 GTAAAGTGATGTGATGACAGGTCAGAGGTGAGATCCACCCCTGGAGTTGGAGACACAGCA 2606  
Db 2392 GTGACCGTGAAGGCGGTGACACGCTTTGAATCCACCCCGGTGTTGGAGACCAATGCA 2451  
Qy 2607 AGATTTAGAGTGGAACACACACCAAAACCAATAGCGGGAATGTGACAAAAAGAAAG 2666  
Db 2452 AGATT-----GGCCACGCTACACTAGCGGTGACTGTGTCCTCAGAGAC 2496  
Qy 2667 CCCCACATCTGATTTCCACTGGAAGCGCAGATGACAAAGAAAGAAAGAAATCACAGGG 2726  
Db 2497 CTTTCACCGCTGATCGTTCCCGCAGAAAGCCACTTGCACAAAGAAAGAGAGAGTGACAGG 2556  
Qy 2727 AAAGAAAGAGACAGTAGAATGGAGGAAATGCTGAAATCACATAGGCGTTACTCAA 2786  
Db 2557 GCAGAA-----GGCAATGCTACCTGTAAAGAG 2586  
Qy 2787 GTGTTACTTGAAGAAAGCTGCAGCAATPACAGATAGTTACTTTGGGCTTTTTCGCATGG 2846  
Db 2587 TTAGTGCTTGGCAGACGCTTGCAGC-----AGAATTATCCAGGCTTTTTCGCCCTGG 2637  
Qy 2847 GAGAAAAAAGTATTTCCAGATCTTCTCGACGAAAGAGTGCATTTGAAGACACAATTC 2906  
Db 2638 GAGAAAAAAGTATTTCCAGACCTTCTTGTAGGAGAGAGTCAATTTGAAGACCCAGTTG 2697  
Qy 2907 GCATACCTTCACTGATGACAAATACTCGGAGGCAACTAAAGATACATTTGCGAGATTC 2966  
Db 2698 GCGTACTTTACAGCCGCAACATACCGGAGGCACTTAAAGATACATTTGCGAGATCC 2757  
Qy 2967 CTCAGATATGTAATPAAATTTCTAAATAGCAAGTTTGGATTCAATCCGCGAAAGTCCCT 3026  
Db 2758 CTTCCGATACCTCAATPAAATTTCTCAACAGCAAGTTTGGATTCAATCCAGGAAAGTCCCT 2817  
Qy 3027 GCTCAGTCCTCAGATGACCGGATTTGATGCAAGAACTGCAAGATATGTCCT 3086  
Db 2818 GCACATGCTCCGCAATGATGACAGATCGTTTATGCAAGAACTCCCAAGATATGTTCCCT 2877  
Qy 3087 GAAGAATTTGACAGACGCTCATTTCAAAAGTGGGCCATTTCTGAGGATPACAGTTTGC 3146  
Db 2878 GAAGAATTTGACAGACTTCATTTCAAAAGTGGGTGCTCACTCTGAGGACATGCAAGTTGCC 2937  
Qy 3147 TTCTCTTATTTTATATCTATGATGAGTGCAGTGCAGCCACTGAAATATATCTCAAGTCTTT 3206  
Db 2938 TTCTCTCTATTTATTTACTCTATGATGAGTGCAGTTCAGCCCTCAATATTTCCCAAGTCTTT 2997  
Qy 3207 GATGAAGTTGACAGATCAATCTGGTGTCTTGTCTGACAGAGAAATCCGAAACACTGGCT 3266  
Db 2998 CATGAAGTAGACAGACCAATCTGCTGTCTGATAGGAAATCCGAAACACTGGCC 3057  
Qy 3267 ACCAGAAATTCAGAACTGCGGTTAAGTTTGAAGATTTGACAGGTTGAGAACACATGCTA 3326  
Db 3058 ACAGAAATTCAGACCTACCTTTAAGCTTGCAGGATTTGACAGGTTTGGAAACACATGTTA 3117  
Qy 3327 ATAAATGCTCAAAATGCTTCCCTGCTGATATCAGCAGCTAAATAATTTCCACCACT 3386  
Db 3118 ATAAATGCTCAAAATGCTTCCCGCTAATATCACTCAACTCAACACATCCACCGACT 3177  
Qy 3387 CAGGAATCTACTATGATCCCAACCTGCCCGTCACTAAAGTCTAGTAACAAACTGT 3446  
Db 3178 CAGGAAGCATACTACGACCCCAACCTGCTCCGGTCACTAAGAGTCTTGTCCACCACTGT 3237  
Qy 3447 AAACAGTAACTCAGAAATCCAAAGCATATAGGACAAACAAATATAGTTTGA 3506  
Db 3238 AAGCCAGTAACTCAGAAATCCAAAGCCTATAAAGACAAAGAAATACAGGTTTGA 3297

Qy 3507 ATCATGGAGAGAGAAATCGCTTTTAAATGATTCTGTACCAACGTTTCTCATGTGGTT 3566  
Db 3298 ATCATGGAGAGAGAAATCGCTTTCAAGATGATACGAAACCAATGTTTCTCATGTGGTT 3357  
Qy 3567 GGCAGTTGGATGACATAAGAAAAAACCCCTAGGAAGTTTGTTCCTGTAATGACAACT 3626  
Db 3358 GGTCAAGTTGGATGACATCAGAAAAAACCCAGGAAGTTTGTTCCTGTAATGACAACT 3417  
Qy 3627 GACCAATCATAAAGATGCTCAGACAGTGAAGGCTTCTCAGGACATCTTATGAATCC 3686  
Db 3418 GACCAATCATAAAGATGCTCAGGACAGTGAAGGCTTCTCAGGACATCTTATGATCC 3477  
Qy 3687 ATGTTCCCATACCTTCCCAATTTGAACTGCCAAAGAGAGTATCGAAACCGTTTCTTCTAT 3746  
Db 3478 ATGTTCCCATACCTTCCCAATTTGAGCTGCCAAAGAGAGTATCGGAACCGCTTCTGAC 3537  
Qy 3747 ATCATGAGCTCAGGAATGGAGGCTTATCGAGACAAATGAGTTTGGACCCATCT 3806  
Db 3538 ATCATGAGCTCAGGAATGGAGGCTTATCGAGACAACTGAGTTTGGACCCACTGC 3597  
Qy 3807 GTACTAGCAACATTTGATTTACTATATTTCTCATTTTCTCATTTTCTGCTGAGCAGTTAA 3866  
Db 3598 GTACTAGCAACATTTGATTTACTATATTTCTCATTTTCTCATTTTCTGCTGAGCAGTTAA 3657  
Qy 3867 CTTAAGCGGAAGATATTTCCAGAGAGGAGATACACAAAGAGCTAGTCCCAATCGAATC 3926  
Db 3658 CTTAAGCGGAAGATATTTCCAGAGAGGAGATACACAAAGAGCTAGTCCAGACCGAATC 3717  
Qy 3927 AGAGTATGAGATCTTCAATTTGAAACCATCTACCTCAGCAGATTTACTGAGCATTTTAA 3986  
Db 3718 AGGTTGAGAGATCTTCAATTTGAAAGTCACTTACCTTAGCATCT--GTGAACTATCCCT 3776  
Qy 3987 ACTCAGCTTTCACAGAGATGTCTTTGTGATGTGCTTA--GCAGTTTGGCCGGAAGAAG 4045  
Db 3777 CCTCGACACACAGCGAGTCCCTGTGATGTGCAAGAGGAGCCTCGTGGGGAAGAG 3836  
Qy 4046 AATATATCAGTACATGCTGTTTGTGCAATGATATAGCCACTGACTAGGAATATT 4105  
Db 3837 GACATCGTCAGACCGGTTCTTCTGCAATGGAAGAGAGCCACTGACTGGAATATT 3896  
Qy 4106 TAACCAACCACTGAAACCTTGTGTGTCAGAGCTCTGAACTGATTTTACTTTTAAAGA 4165  
Db 3897 ----CAGCACATAGAACCTGTGCTAATAGC---TTGTACAGCTTGTACTTTAAAG 3948  
Qy 4166 ATTTGCTATGACCTGTCTATCTTTTATPAAAGGCTCATGACAGAGACAGCTGT 4225  
Db 3949 ATTTGCGAAGGACCTGTGGCTTGTGACAAACCTCCTCCTGACA-----AGTGTCT 4000  
Qy 4226 AATTTCCACAGCAATCATTTGACAGCTAATCTTTATGAGAGAGCCCTATGCCAGCTGGGA 4285  
Db 4001 GGTTCCTTCCCAGTTACTGAGACT-----GAGAAACAGTCCATCTTGAA 4049  
Qy 4286 GTCATGCTAAGAGGCTCCAGTCTTTGCAATCCAAAGCCTTTTGTGTAAGTTTGCAT 4345  
Db 4050 GCAAGTGGAGGGGCCAGTCTTTGCAATCCAAAGCTTTCCAGCATAAATTTCTGGCT 4109  
Qy 4346 TT-----TTTTTTTCAATTTCCATTTT-----AAGTAGTTACTAGTTAACT 4388  
Db 4110 GTCCTCTCTTTGATCCATTTTCCATTTT-----TAAACAAATAGTGGCTACTAGTT 4169  
Qy 4389 AGTTATTTCTGCTTCTGAGTATAACGAATTTGGGATGCTCTAAACCTATTTTATAGATGT 4448  
Db 4170 AGTCATTTCTCATTCTCAAAATAACAAATCAGGATGTCAAAAC--ATTGTATAGATCT 4227  
Qy 4449 ATTTAAATATGAGCAATATCACCTCTTATGCAATACCTTAAATATGAGTTTATTA 4508  
Db 4228 ATTTAAATATATAGAACGATTTACTTCT--TTAGCCTATCTAAATATGATTTTATTA 4285  
Qy 4509 ATATTTAAGACTGTAATGCTTAAACCACT---AACTACTGAGAGCTCAATGATGCA 4565  
Db 4286 ACAGTCAAG-----TGGTCTTGAACCGCTAACACTACTGAGAGGCTCGA--GATTGA 4336







QY	3447	AAA	CCAGTAACTGACAA	AAATCCACA	AAAGCATAT	TAAGGACAAA	AAACAAATAT	PAGTGT	TGAA	3506									
DB	3238	AAGCCAGTAACTGAC	AAAGATCCACA	AAAGCCTAT	TAAGACAA	GAAGAA	CAAATACAGG	TTTTGAA		3297									
QY	3507	ATCATGGGAGAA	GAAATCGCTTTT	TAAATGATTCGT	ATCCAA	CGTTTCTCAT	GCTGCTTT	CTCATGCTG	TTT	3566									
DB	3298	ATCATGGGAGAA	GAAATCGCTTT	CAAGATGATA	GAACCA	TGTTTCTCAT	GCTGCTTT	CTCATGCTG	TTT	3357									
QY	3567	GGCCAGTTGGAT	GTACATAA	GAAAAAA	AAACCTTAGG	AAAGTTTGT	TGCTTGA	ATGACAA	CATT	3626									
DB	3358	GGTCAGTTGGAT	GTACATCAG	AAAAAA	CCCCAGG	AAAGTTTGT	TGCTTGA	ATGACAA	CATT	3417									
QY	3627	GACCACAA	TATAAAGATGCT	CACAGAGT	GAAGGCTGTCT	CAGGAGCTTCT	TATGAT	GTCC		3686									
DB	3418	GACCACAA	CCCAAGATGCCCCG	ACAGTGAAGGCTGTCT	CAGGAGCTTCT	TATGAT	GTCC			3477									
QY	3687	ATGTTTCCCCAT	PACCTTCCCA	ATTGAACTG	CCAAAGAGAGAT	ATCGAA	ACCGTTTCC	TTTCAT		3746									
DB	3478	ATGTTTCCCCAT	PACCTTCCCA	ATTGAGCTG	CCAAAGAGAT	ATCGAA	ACCGTTTCC	TTTCAC		3537									
QY	3747	ATGCATGAGCTG	CAGGAATGGAGGGCT	TATCGACACAA	ATGAAGTTTTCG	AGCCATTGT				3806									
DB	3538	ATGCATGAGCTG	CAAGAAATGGCGGC	ATATCGAGACAA	AGCTGAAGTTTTCG	AGCCACTGC				3597									
QY	3807	GTA	CTAGCAACAT	TGATTATGTTT	PACTATATTTCT	CATTTTTGCT	GAGCAGTTAA	TTGCA		3866									
DB	3598	GTA	CTAGCAACGTTG	ATTATTTACTAT	TATCTCATTTTTT	TGCTGA	CAGATA	TTTGCT		3657									
QY	3867	CTTAA	CGGAAAGATATTTCC	CCAGAGAGAGAT	ATACAA	GAAGCTAGTCC	CAATCG	ATC		3926									
DB	3658	CTGAAGCG	AAAGATATTTCC	CAGAGAGAGAT	ATACAA	GAAGCTAGTCC	CAAGAC	CGAATC		3717									
QY	3927	AGAGTATAG	AAAGATCTTCATTTG	AAAAACCATCTAC	CTCAGCATTTACT	GAGCATTTTAA				3986									
DB	3718	AGGGTGTAG	AAAGATCTTCATTTG	AAAGTACCTTAC	CTTAGCATCT	GTGAACATCT	CCCT			3776									
QY	3987	ACTCAGCTT	CACAGAGATGCTTTT	TGTGATGTGAT	GTCTTA	GCAGTTTGG	CCCCGAA	GAAGG		4045									
DB	3777	CCTCGAC	ACACACAGCGGAGTCC	CTGTGATGTGG	CACAGAGG	CAGCCTCGTGG	GGGAA	GAAGG		3836									
QY	4046	AAA	ATATCCAGTACCA	TGCTGTTTGT	TGGCATGATAT	TAGCCCACTG	ACTAG	AGAA	TATT	4105									
DB	3837	GACATCGT	GCAGACGGGTCTCT	CTGCAATGG	GAAGAGAGCCCA	CTGACCTG	AGAA	TATT		3896									
QY	4106	TAA	CCAAACCCACTG	AAAACTTGTGTG	TCGAGCAGCTCTG	AACTGATTTTACT	TTTTTAA	AGA		4165									
DB	3897	-----	CAGCACAT	TAAGAACCTGTGT	CAATAGC-----	TTGTACAGCTTGT	TACTTTTTA	AGG		3948									
QY	4166	ATT	TGCTCATG	GAGCTGCTGATCC	TTTTTATA	AAAGGCTCA	CTGAC	AA	GAGACAGCTGTT	4225									
DB	3949	ATT	TGCGAGAGACCTGT	CGGCTTGTTTG	ACAAAACCCCTCC	CTGACA-----	AGCTGCT			4000									
QY	4226	AA	TTTCCACAGCA	ATCATTTGC	AGACTAACTTTT	TATTAGG	GAAGACCTAT	GCCAGCT	GGGA	4285									
DB	4001	GG	TTCTTCC	CCAGTTACT	GCAGACT-----	-----	GAGAA	AC	CAGTCCATCTT	GA	4049								
QY	4286	GT	GA	NTGCTAAGAGGCT	CCAGTCTTTG	CAATCCAA	AGCCTTTTGT	CTAA	AGTTTGT	GC	4345								
DB	4050	GCA	AGTGC	GGAGGGGCCCA	GTCTTG	CAATCCAA	AGCTTTCC	CAGCAT	TAATTTCT	GGC	4109								
QY	4346	TT-----	TTTTTTT	TCATTTCC	CACTTTT-----	-----	AACTAG	T	CTA	AGTTAA	4388								
DB	4110	GT	CTCCTCCTTT	GTATCCATTTTCC	CAATTTTTTTTTT	TTTTTAAAA	AA	CA	TA	AGTGGCTACT	AA	4169							
QY	4389	AG	TATTCTT	GCTTCTG	AGTATAACGA	NTTGGATGTCT	TA	AA	ACCTATTTTTT	TATAG	AT	4448							
DB	4170	AG	T	CA	TCTCACTTCT	CAAAATAACAA	ATCAGG	AT	GTCA	AAAC	-	ATT	GT	AT	AG	AT	CT		4227
QY	4449	ATT	TAA	TATATG	CGCA	NTATCA	CCCTCTTAT	TG	CA	CA	ATAC	CT	TA	AA	TATAG	TTTT	TATTA		4508
DB	4228	ATT	TAA	ATA	TATAG	AAACG	ATTA	CTCTCT	TTAG	CC	CTATCT	CA	AA	TATG	ATTT	TATTTATTA		4285	
QY	4509	AT	ATTT	TAAG	ACTGT	TAAT	TGCTTATA	ACCA	CT---	AACT	ACT	GA	AG	AG	CTCA	AT	GA	TTCA	4565

Db	4286	ACAGCTCAAG-----TGGTCTTTGAACCGCTAACCACTACTGAAGAGCTCGA-GAATTGA	4336
Qy	4566	CATCTGAAATGCTTTGTAATTATTGACTTTCAGGCCCTTAAGAATGCTATGATTTCAAGTGC	4625
Ds	4337	CGTTGAAGTGCCTTGAGCTTGTTAACTCATTCGCCAGAATACTGTGACCTCGTGTGC	4396
Qy	4626	AGGCTCAATTTCAACAGGCTAGAGT-----TAGTACTACTTACCAGATGTHAATTATGT	4678
Ds	4397	GGGCTCATTTGCCGAAGGGCTAGTGTCAAGTAGCAGTGTCTCTCACCGGATGTAATTATGT	4456
Qy	4679	TTTCGAAATGTACATATTTCAAAACAGAGTGCCTCATTTTGAAGATGAGTAGTCTGATGG	4738
Ds	4457	CGTGGAAATGTACATA--CAGACAAAGTGCCTCACTTCAGAAATGAGTAGTCTGATGG	4514
Qy	4739	CATCGGCACATTACAGTG	4756
Ds	4515	CACCAGCGAGTGATGGTG	4532
RESULT 15			
AD	AD62652		
ID	AD62652	standard; DNA; 5229 BP.	
XX	XX	AC	AD62652;
XX	XX	AC	AD62652;
DT	DT	15-JAN-2004	(first entry)
XX	XX	Mouse DNA #1 used to illustrate the method of the invention.	
XX	XX	Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;	
KW	KW	N-acetylglucosamine-1-phosphotransferase; gene therapy; mouse; ds.	
XX	XX	Mus musculus.	
OS	OS	Mus musculus.	
XX	XX	US2003124653-A1.	
PN	PN	03-JUL-2003.	
PD	PD	21-DEC-2001; 2001US-00023890.	
XX	XX	21-DEC-2001; 2001US-00023890.	
PR	PR	(NOVA-) NOVAZYME PHARM INC.	
XX	XX	Canfield WM;	
PI	PI	WPI; 2003-810985/76.	
XX	XX	Producing a glycoprotein with reduced complex carbohydrates by culturing	
DR	DR	the lectin resistant mammalian cell expressing the glycoprotein for	
PT	PT	treating lysosomal storage disease.	
XX	XX	Disclosure; Page 20-23; 46pp; English.	
XX	XX	The present invention provides a method of producing a glycoprotein	
CC	CC	having reduced complex carbohydrates by culturing the lectin resistant	
CC	CC	mammalian cell expressing the glycoprotein. The method is useful for	
CC	CC	producing a glycoprotein with reduced complex carbohydrates for treating	
CC	CC	lysosomal storage disease. The present invention is also useful in gene	
CC	CC	therapy. The present sequence is mouse DNA used to illustrate the method	
CC	CC	of the invention	
XX	XX	Sequence 5229 BP; 1485 A; 1242 C; 1199 G; 1302 T; 0 U; 1 Other;	
SQ	SQ	Query Match 50.4%; Score 2820.8; DB 10; Length 5229;	
		Best Local Similarity 79.1%; Pred. No. 0;	
		Matches 3667; Conservative 1; Mismatches 833; Indels 137; Gaps 21;	
Qy	150	GGCGGTGAAGGGGTGATGTCTTCAAGCTTCCTGCAGAGCAACCTATACCTGCTGTCC	209
Ds	1	GGCGGTGAAGGGGTGATGTCTTCAAGCTTCCTGCAGAGCAACCTATACCTGCTGTCC	60



Db 2221 CTGCTAAGTCTTTCTCGGGAAATTCAGTAGATATAAAATAAACCTCAAGCT---AGG 2277  
Qy 2427 ACAGATGAACAAATGACAGTTTGGTGGCTCCACAGGAAAAACAGGTTTCATAAAAGCATC 2486  
Db 2278 ACCGATGAACAAAGGCAACCTGGAGTCCACAGGAARACCTTCTCAGAGC----- 2333  
Qy 2487 TTGCCAAACAGCTTGAAGAGTGTGAAAGATGCGAGGTTGACTTTTCTTCGAGTGAGT 2546  
Db 2334 --TCCACATGGCTTTGCTGGTGAAACACAGATCAGAGAGATGCACTGGCCCCAGCAGAGACA 2391  
Qy 2547 GTAAAGTGAATGTCATGACCGAGGTGAGAAATCCACCCCTGGACTTGGAGACACAGCA 2606  
Db 2392 GTGACCGTGAAGGCGGTGACCACTTTGAAATCCACCCCGGTGTTGGAGACCAATGCA 2451  
Qy 2607 AGATTTAGAGTGAATCTCACACCCAAAAACCAATAGCGGAAATGTGACAAAAAGAAAG 2666  
Db 2452 AGATT-----GGCCCGAGCTACACTAGGCGTGACTGTGTCCAAAGAGAAC 2496  
Qy 2667 CCCCATCTCTGATTGTTCCACTGGAAGCCAGATGACAAAAGAAAGAAATCACAGGG 2726  
Db 2497 CTTTCACGCTGATCGTTCCCCAGAAAGCACTTGCACAAAGAGAGAGTGACAGG 2556  
Qy 2727 AAGAAAAAGAGAACAGTAGAATGGAGAAATGCTGAATATCATAGGCGTTACTGAA 2786  
Db 2557 GCAGAA-----GGCAATGCTGTACCTGTAAAGGAG 2586  
Qy 2787 GTGTTACTTGAAGAAAGCTGAGCATTAACAGATAGTTACTTGGGCTTTTGGCCATGG 2846  
Db 2587 TTATGCTGCTGAGAGCGTTGCGAGC-----AGAAATACAGGCTTTTGGCCCTGG 2637  
Qy 2847 GAGAAAAAAGTATTTCCAGATCTCTCCAGCAAGAGAGTCAATGAGAGACAAATG 2906  
Db 2638 GAGAAAAAAGTATTTCCAGACCTTCTGATGAGAGAGATCATTTGAGACCCAGTTG 2697  
Qy 2907 GCATCTTCACTGATGACAAAAATACCTGGAGGCAACTAAAGATAATTTGCAAGATCC 2966  
Db 2698 GCGTACTTTACAGACCGCAACATACCGGAGGCAACTAAAGATACATTTGACAGATCC 2757  
Qy 2967 CTCAGATGATAAATAAATTTCTAAATAGCAAGTTTGGATTACATCGCGGAAGTCCCT 3026  
Db 2758 CTCGATACGTCATAAATAATTTCTCAACAGCAAGTTTGGATTACATCCAGAAAGTCCCT 2817  
Qy 3027 GCTCACATGCTCATGATGACCGGATTTGATGCAAGAACTGCAAGATATGTTCCCT 3086  
Db 2818 GCACACATGCGGCACATGATTGACAGGATCGTTATGCAAGAACTCCAAAGATATGTTCCCT 2877  
Qy 3087 GAAGAAATTTGACNAGACGTCATTTCAAAAGTGGCCATCTGAGGATATGCAAGTTGCC 3146  
Db 2878 GAAGAAATTTGACAGAGACTTCAATTTCAAAAGTGGCTCACTCTGAGGACATGCAAGTTGCC 2937  
Qy 3147 TTCTCTTATTTTATTTATCTCATGAGTGCAGTGAGGCACTGAATATATCTCAAGTCTTT 3206  
Db 2938 TTCTCTTATTTTATTTATCTCATGAGTGCAGTTGAGCCCTCATATATTTCCAGTCTTT 2997  
Qy 3207 GATGAAGTTGATACAGATCAATCTGGTGTCTGTCAGAGAGAAATCCGAACACTGGCT 3266  
Db 2998 CATGAAGTAGACACAGACCAATCTGGTGTCTGTCATAGGGAATCCGAACACTGGCC 3057  
Qy 3267 ACCGAATTCAGAACTGCGGTAAAGTTTGGAGATTTGACAGTCTGGAACACATGCTA 3326  
Db 3058 ACGGAATTCAGACCTTACCTTTAAGCTTGCAGGATTTGACAGTTTGGAAACATGTTA 3117  
Qy 3327 ATAAATTTGCTCAAAAATGCTCCCGCTAAATATCACTCAACTCAACAAACATCCACCGACT 3386  
Db 3118 ATAAATTTGCTCAAAAATGCTCCCGCTAAATATCACTCAACTCAACAAACATCCACCGACT 3177  
Qy 3387 CAGGAATCTCTACTATGATCCCAACCTGCCAGGTCATAAAGTCTAGTAACAAACTGT 3446  
Db 3178 CAGGAATCTACTAGGACCCCAACCTGCTCCGCTCACTAAGAGTCTTTGTCACCAACTGT 3237  
Qy 3447 AAACAGTAACTGACAAATCCAAAGCATATTAAGGACAAAAACAAATATAGGTTTGA 3506  
Db 3238 AAGCCAGTAACTGACAAATCCAAAGCATATTAAGGACAAAAACAAATATAGGTTTGA 3297

Qy 3507 ATCATGGGAGAGAGAAATCGCTTTTAAATGATTCTGTACCAAGTTTCTCATGTGGTT 3566  
Db 3298 ATCATGGGAGAGAGAAATCGCTTTCAAGATGATACGAACCAATGTTTCTCATGTGGTT 3357  
Qy 3567 GGCAGTTGGATGACATAAAGAAAAACCTAGAGAGTTTGTGCTGAATGACAAATTT 3626  
Db 3358 GGTCAAGTTGGATGACATCAGAAAAAACCCAGAGAGTTGCTTTGTCTGAATGACAAATTT 3417  
Qy 3627 GACCAAAATCATAAAGATGCTCAGACAGTGAAGGTGTTCTCAGGACTTCTATCAATCC 3686  
Db 3418 GACCAAAATCATAAAGATGCTCAGAGAGTCCCGGACAGTGAAGGTGCTCTCAGGACTTCTATGAGTCC 3477  
Qy 3687 ATGTTCCCATACCTTCCCAATTTGAATGTCGAAGAGTATCGAAAAACGTTTCCTTCAAT 3746  
Db 3478 ATGTTTCCCATACCTTCCCAATTTGAATGTCGAAGAGTATCGAAAAACGTTTCTGCAC 3537  
Qy 3747 ATGTCATGCTCGAGAAATGAGGCTTATCGAGACAAATGGAAGTTTGGACCCCATGTT 3806  
Db 3538 ATGTCATGCTCGAGAAATGAGGCTTATCGAGACAAATGGAAGTTTGGACCCCATGTT 3597  
Qy 3807 GTACTAGCAACATTTGATTATGTTTACTATATTTCTCATTTTGTGTCAGCAGTTAATGCA 3866  
Db 3598 GTACTAGCAACATTTGATTATATTTACTATATTTCTCATTTTGTGTCAGCAATAATGCT 3657  
Qy 3867 CTTAAGCGGAAGATATTTCCAGAGAGGAGTATACAAAGAGCTAGTCCCAATCGAATC 3926  
Db 3658 CTGAAGCGGAAGATATTTCCAGAGGAGGATATACAAAGAGCTAGTCCCAAGCGAATC 3717  
Qy 3927 AGAGTATAGAAATCTTCTCATTTGAAAAACCATCTACTCTCAGCATTTTACTGAGCATTTTAAA 3986  
Db 3718 AGGTTGTAAGATCTTCTCATTTGAAAGTCACTACTCTTAGCATCT-GTGAACATCTCCCT 3776  
Qy 3987 ACTAGCTTCAAGAGATGCTTTGTGATGATGATGCTTA-GCAGTTTGGCCCGGAAGAGG 4045  
Db 3777 CCTCGACACACACAGGAGTCCCTGTGATGTGGCAAGAGGAGGCTCTGTTGGGAGAGG 3836  
Qy 4046 AAAATATCCAGTACCATGCTGTTTGTGGCATGAATATAGCCCACTGACTAGGAATATT 4105  
Db 3837 GACATCGTCAGACCGGGTTCTTCTGCAATGGAGAGAGGCCACTGACTTGGAAATATT 3896  
Qy 4106 TAAACAAACCACTGAAAACTTGTGTGTCAGCAGCTCTGAACTGATTTTACTTTTAAAGA 4165  
Db 3897 ----CAGCACACTAAGAACCTGTGTCAATAGC----TTGTACAGCTTGTACTTTTAAAGG 3948  
Qy 4166 ATTTGCTCATGGACCTGTCATCTTTTATAAAAGGCTCACTGACAGAGACAGCTGTT 4225  
Db 3949 ATTTGCGAGAGACCTGCGGCTTGTGACAAACCCCTCCCTGACA-----AGCTGCT 4000  
Qy 4226 AATTTCCCAAGCAATCATTTGACAGTAACTTTATTAGGAGAGGCTTATGCGAGCTGGGA 4285  
Db 4001 GGTTTCTTCCCGAGTACTGAGACT-----GAGAAACCACTGCTCATCTTGAAA 4049  
Qy 4286 GTGATGTGTCAGAGCTCCAGTCTTGTGCAATCCAAAGCCTTTTGTAAAGTTTGTGCACTT 4345  
Db 4050 GCAAGTGGGAGGCGGCCAGTCTTTGCAATTTCCAAAGCTTTCCAGCATAAATTTCTGGGCTT 4109  
Qy 4346 TT-----TTTTTTTCAATTTCCCAATTTT-----AAGTAGTTTACTAAGTTAACT 4388  
Db 4110 GTCTCTCTCTTTGATCCATTTCCCATTTT-----TAAAAAAACATTAAGTGGCTACTAAGTT 4169  
Qy 4389 AGTTATCTTGTGCTGTCAGTATACGAATGCGGATGCTTAAACCTATTTTATAGATGTT 4448  
Db 4170 AGTCAATCTCACTTCTCAAAATAACAAATCAGGATGTCAAAAAC--ATTGTATAGATCTT 4227  
Qy 4449 ATTTAAATAATGACAGCAATATCACCTCTTATTGCAATATCTAAATATAGTTTATTA 4508  
Db 4228 ATTTAAATAATGAGACGATTAATCTTCT--TTAGCTATCTAAATATGATTTTATTA 4285  
Qy 4509 ATATTTAAGACTGTAAATGGTCTTAAACCACT---AACTACTGAGAGCTCAATGATTGA 4565  
Db 4286 ACAGTCAAG-----TGGTCTTGAACCGCTTAAACCACTACTGAGAGCTCGA-GATTGA 4336

QY	4566	CATCTGAATGCTTTGTAAATTATTGACTTCAGCCCTAAGAATGCTATGATTTCACGTGC	4625
Db	4337	CGTTGAAGTGTCTTTGAGCTTGTAACTCATTCCTCCCAAGATACTGTGACCTCGTGTGC	4396
QY	4626	AGGTCTAAATTTCAACAGGCTAGAGT-----TAGTACTTACCAGATGTAATTATGT	4678
Db	4397	GGGCTGATTGCGAAGGGCTAGTGTACGTAGCAGTGTCTCACCCGATGTAATTATGT	4456
QY	4679	TTTGGAATGTACATATTCAACAGAGTGCCTCATTTTAGAAATGAGTGTGCTGATGG	4738
Db	4457	CGTGAATGTACATA--CAGCAAAAGTGCCTCACTTCAGAAATGAGTGTGCTGATGG	4514
QY	4739	CACTGCGACATTACAGTG	4756
Db	4515	CACCAGCGAGTGATGGTG	4532

Search completed: November 21, 2004, 23:56:11  
Job time : 2432 secs

***This Page Blank (uspto)***







QY 2581 CACCCCTGGACTTGAGACCACAGCAAGATTTAGAGTGGAAACTCAGACCCCAAAAAACCA 2640  
Db 2581 CACCCCTGGACTTGAGACCACAGCAAGATTTAGAGTGGAAACTCAGACCCCAAAAAACCA 2640  
QY 2641 TAGCGGAAATGTGCAAAAGAAAAGCCCCCATCTCTGATTTGTTCCACTGGAAGCCAGA 2700  
Db 2641 TAGCGGAAATGTGCAAAAGAAAAGCCCCCATCTCTGATTTGTTCCACTGGAAGCCAGA 2700  
QY 2701 TGACAAAAGAAAAGAAAATCAGCGGAAAGAAAAGAACAGATAGATGAGGAAAATG 2760  
Db 2701 TGACAAAAGAAAAGAAAATCAGCGGAAAGAAAAGAACAGATAGATGAGGAAAATG 2760  
QY 2761 CTGAAAATCACATAGGCGTTACTGAAAGTGTACTTTGGAAGAAAGCTGCGAGCATTACACAG 2820  
Db 2761 CTGAAAATCACATAGGCGTTACTGAAAGTGTACTTTGGAAGAAAGCTGCGAGCATTACACAG 2820  
QY 2821 ATAGTTACTTGGGCTTTTGGCATGGGAGAAAAGAAAGTATTTCCAGATCTTCTCGACG 2880  
Db 2821 ATAGTTACTTGGGCTTTTGGCATGGGAGAAAAGAAAGTATTTCCAGATCTTCTCGACG 2880  
QY 2881 AAGAAAGAGTCATTGGAAGACACAATTTGGCATCTTCATCTGATAGCAAAAATACTGGAGGC 2940  
Db 2881 AAGAAAGAGTCATTGGAAGACACAATTTGGCATCTTCATCTGATAGCAAAAATACTGGAGGC 2940  
QY 2941 AACTAAAAGATACATTTGCGAGATCCCTCAGATATGTAAATTAATAATAGCAAGT 3000  
Db 2941 AACTAAAAGATACATTTGCGAGATCCCTCAGATATGTAAATTAATAATAGCAAGT 3000  
QY 3001 TTGGATTCACATCGCGGAAAGTCCCTGCTCAGATGCTCAGATGATGACCGGATGTGTA 3060  
Db 3001 TTGGATTCACATCGCGGAAAGTCCCTGCTCAGATGCTCAGATGATGACCGGATGTGTA 3060  
QY 3061 TGCAAGAACTCAAGATATGTTCCCTGAAGAAATTTGCAAGACGTCATTTTCAAAAATGTC 3120  
Db 3061 TGCAAGAACTCAAGATATGTTCCCTGAAGAAATTTGCAAGACGTCATTTTCAAAAATGTC 3120  
QY 3121 GCAATTCGAGATATGAGTTGCGCTCTCTATTTTATATCTCATGATGCGAGTGC 3180  
Db 3121 GCAATTCGAGATATGAGTTGCGCTCTCTATTTTATATCTCATGATGCGAGTGC 3180  
QY 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTCTTGT 3240  
Db 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTCTTGT 3240  
QY 3241 CTGACAGAAATCCGAACACTGGCTACAGAAATTCAGAACTGCGGTTAGTTGCGAG 3300  
Db 3241 CTGACAGAAATCCGAACACTGGCTACAGAAATTCAGAACTGCGGTTAGTTGCGAG 3300  
QY 3301 ATTTGACAGCTCTGGAACACATGCTAATAAATTTGCTCAAAAATGCTTCTGCTGATATCA 3360  
Db 3301 ATTTGACAGCTCTGGAACACATGCTAATAAATTTGCTCAAAAATGCTTCTGCTGATATCA 3360  
QY 3361 CGCAGCTAAATATTTCCAACTCAGGAACTCTATGATCCCAACCTGCGACCGG 3420  
Db 3361 CGCAGCTAAATATTTCCAACTCAGGAACTCTATGATCCCAACCTGCGACCGG 3420  
QY 3421 TCACATAAAGTCTGATACAACTCTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAA 3480  
Db 3421 TCACATAAAGTCTGATACAACTCTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAA 3480  
QY 3481 AGGACAAAACAAAATATAGTTTGAATCATGGGAGAAAGAAATCGCTTTTAAATGA 3540  
Db 3481 AGGACAAAACAAAATATAGTTTGAATCATGGGAGAAAGAAATCGCTTTTAAATGA 3540  
QY 3541 TTGCTACCAACGGTTCTCATGTGTTGGCAGTTGGATGACATAGAAAACCTTAGGA 3600  
Db 3541 TTGCTACCAACGGTTCTCATGTGTTGGCAGTTGGATGACATAGAAAACCTTAGGA 3600  
QY 3601 AGTTTGTTCCTGGAATGACAACTTGACCAATCATAGATGCTCAGACAGTGAAG 3660  
Db 3601 AGTTTGTTCCTGGAATGACAACTTGACCAATCATAGATGCTCAGACAGTGAAG 3660  
QY 3661 CTGTTCTCAGGACTTCTATGAATCAATGTTCCCATACCTTTCCCAATTTGAATGCGCAA 3720

Db 3661 CTGTTCTCAGGACTTCTATGAATCAATGTTCCCATACCTTCCCAATTTGAATGCGCAA 3720  
QY 3721 GAGAGTATCGAAACCGTTTCCCTTCATATGATGAGTGCAGAAATCGAGGCTTATCGAG 3780  
Db 3721 GAGAGTATCGAAACCGTTTCCCTTCATATGATGAGTGCAGAAATCGAGGCTTATCGAG 3780  
QY 3781 ACAAATTTGAAGTTTGGACCCATTGTGTACTAGCAACATTGATTAATGTTTACTATATCT 3840  
Db 3781 ACAAATTTGAAGTTTGGACCCATTGTGTACTAGCAACATTGATTAATGTTTACTATATCT 3840  
QY 3841 CATTTTGTGCTGAGAGTTAAATTTGCACTTAAGCGGAGATATTTCCCAAGAGGAGATAC 3900  
Db 3841 CATTTTGTGCTGAGAGTTAAATTTGCACTTAAGCGGAGATATTTCCCAAGAGGAGATAC 3900  
QY 3901 ACAAAGAGCTAGTCCCAATCGAATCAGAGTATAGAAATCTTCATTTGAAAACCACTCTA 3960  
Db 3901 ACAAAGAGCTAGTCCCAATCGAATCAGAGTATAGAAATCTTCATTTGAAAACCACTCTA 3960  
QY 3961 CCTCAGCATTTACTGAGCATTTTAAACTCAGGTTTCAAGAGATGCTTTGTGATGTGAT 4020  
Db 3961 CCTCAGCATTTACTGAGCATTTTAAACTCAGGTTTCAAGAGATGCTTTGTGATGTGAT 4020  
QY 4021 GCTTAGCAGTTTGGCCGAGAGGAAATATCCAGTACCAGTCTCTTTTGTGCGCATGAA 4080  
Db 4021 GCTTAGCAGTTTGGCCGAGAGGAAATATCCAGTACCAGTCTCTTTTGTGCGCATGAA 4080  
QY 4081 TATAGCCCACTGACTAGGAATTTAAACCAACCCACTGAAAACCTTTGTGTCGAGCAGC 4140  
Db 4081 TATAGCCCACTGACTAGGAATTTAAACCAACCCACTGAAAACCTTTGTGTCGAGCAGC 4140  
QY 4141 TCTGAACCTGATTTTACTTTTAAAGATTTGCTCATGAGCCTGTCATCTTTTATAAAAA 4200  
Db 4141 TCTGAACCTGATTTTACTTTTAAAGATTTGCTCATGAGCCTGTCATCTTTTATAAAAA 4200  
QY 4201 GGTCTACTGACAGAGACAGCTGTTAATTTCCCAAGCAATCAITGCGAGCTAACTTTAT 4260  
Db 4201 GGTCTACTGACAGAGACAGCTGTTAATTTCCCAAGCAATCAITGCGAGCTAACTTTAT 4260  
QY 4261 TAGGAGAGCCTATGCGAGCTGGAGTGTGTAAGAGGCTCCAGTCTTTGCAATTCCTCAA 4320  
Db 4261 TAGGAGAGCCTATGCGAGCTGGAGTGTGTAAGAGGCTCCAGTCTTTGCAATTCCTCAA 4320  
QY 4321 AGCCTTTGCTAAAGTTTTCACATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCATTT 4380  
Db 4321 AGCCTTTGCTAAAGTTTTCACATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCATTT 4380  
QY 4381 AGTTAACTAGTTATTTCTGCTTCTGATATAACGAATTTGGGATGCTTAAACCTTATTTTA 4440  
Db 4381 AGTTAACTAGTTATTTCTGCTTCTGATATAACGAATTTGGGATGCTTAAACCTTATTTTA 4440  
QY 4441 TAGATGTTTAAATAATGACAGCAATATCACCTCTTATTTGCAATACCTTAATTTATGAG 4500  
Db 4441 TAGATGTTTAAATAATGACAGCAATATCACCTCTTATTTGCAATACCTTAATTTATGAG 4500  
QY 4501 TTTTATTAATTTTAAAGCTGTAATTTGCTTAAACCACTAACTACTGAAAGAGCTCAATG 4560  
Db 4501 TTTTATTAATTTTAAAGCTGTAATTTGCTTAAACCACTAACTACTGAAAGAGCTCAATG 4560  
QY 4561 ATTGACATCTGAAATGCTTTGTAATTTATGACTTTCAGCCCTTAAAGAAATGCTATGATTC 4620  
Db 4561 ATTGACATCTGAAATGCTTTGTAATTTATGACTTTCAGCCCTTAAAGAAATGCTATGATTC 4620  
QY 4621 CGTGCGGCTCTAAATTTCAACAGGCTAGGTTAGTACTTACTTACAGATGTAATTTATGTTT 4680  
Db 4621 CGTGCGGCTCTAAATTTCAACAGGCTAGGTTAGTACTTACTTACAGATGTAATTTATGTTT 4680  
QY 4681 TGGAAATGTACATATTTCAAAAGAGTGCCTATTTTGAAGAAATGAGTAGTGTGATGGCA 4740  
Db 4681 TGGAAATGTACATATTTCAAAAGAGTGCCTATTTTGAAGAAATGAGTAGTGTGATGGCA 4740  
QY 4741 CTGCAATTTACAGTGGTGTCTTGTATTAATCTCATTTGTTATTTCCAGTAGCTACTCT 4800

Db 4741 CTGGCACATTACAGTGGTCTTGTGTTAACTACTCATTTGGTATATTCAGTAGCTATCTCT 4800  
Qy 4801 CTCAGTTGGTTTTTGTATAGAACAGAGGCCAGCAAACTTTCTTTGTAAGAGCTGGTTAGT 4860  
Db 4801 CTCAGTTGGTTTTTGTATAGAACAGAGGCCAGCAAACTTTCTTTGTAAGAGCTGGTTAGT 4860  
Qy 4861 AAATTAATTCAGAGCCACCTGTCTTGTCTATACATCTTCTTCTGCTGTGTTAGTTTGT 4920  
Db 4861 AAATTAATTCAGAGCCACCTGTCTTGTCTATACATCTTCTTCTGCTGTGTTAGTTTGT 4920  
Qy 4921 TTTTCTTCAAAACAACTCTTAAATGTAATAATGTAATAATGTAATAATGTAATAATGTA 4980  
Db 4921 TTTTCTTCAAAACAACTCTTAAATGTAATAATGTAATAATGTAATAATGTAATAATGTA 4980  
Qy 4981 ACTGCCACAGCCAGATGTGACCTCAGGCCATCTTGGCCATCTATGCAATCTAGAGATTTT 5040  
Db 4981 ACTGCCACAGCCAGATGTGACCTCAGGCCATCTTGGCCATCTATGCAATCTAGAGATTTT 5040  
Qy 5041 TTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 5100  
Db 5041 TTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 5100  
Qy 5101 TCGAGTGGCGAACTCAGCTCAGTCACTGCAACCTCCGCTCCGCTTCAAGCAGTTCTGTC 5160  
Db 5101 TCGAGTGGCGAACTCAGCTCAGTCACTGCAACCTCCGCTCCGCTTCAAGCAGTTCTGTC 5160  
Qy 5161 TCAGCTTCTCAGTAGCTGGACCTACAGGTGCATGCCACACACACCTCTGCTAATTTTGTGTA 5220  
Db 5161 TCAGCTTCTCAGTAGCTGGACCTACAGGTGCATGCCACACACACCTCTGCTAATTTTGTGTA 5220  
Qy 5221 TTTTGTAGTAGAGCGGGGTTTCCACCATATTTGGTCAAGCTTATCTTGAACCTCTGACCTC 5280  
Db 5221 TTTTGTAGTAGAGCGGGGTTTCCACCATATTTGGTCAAGCTTATCTTGAACCTCTGACCTC 5280  
Qy 5281 AGGTGATCCAGCTGCTGCTCCCAAGTGCAGATTCAGGATACAGGATACAGGATGAC 5340  
Db 5281 AGGTGATCCAGCTGCTGCTCCCAAGTGCAGATTCAGGATACAGGATACAGGATGAC 5340  
Qy 5341 CCAGCCGAGATTTAGTATTTTATGTTATGTTTAACTCTGGCTGAGCCATTTTATG 5400  
Db 5341 CCAGCCGAGATTTAGTATTTTATGTTATGTTTAACTCTGGCTGAGCCATTTTATG 5400  
Qy 5401 TCATAATACATGATTTGTAAGAGCAGATTCATGAGTAATCTCTGACAGTTATTTAG 5460  
Db 5401 TCATAATACATGATTTGTAAGAGCAGATTCATGAGTAATCTCTGACAGTTATTTAG 5460  
Qy 5461 ATCATGATCTCAAAATATCTCCCAATGGCATACATCTTTTGTACAAAGAACTTGAA 5520  
Db 5461 ATCATGATCTCAAAATATCTCCCAATGGCATACATCTTTTGTACAAAGAACTTGAA 5520  
Qy 5521 ATGTAATACATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 5580  
Db 5521 ATGTAATACATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 5580  
Qy 5581 AAGTTAAATTTTGAATA 5597  
Db 5581 AAGTTAAATTTTGAATA 5597

## RESULT 2

US-09-636-077A-4  
; Sequence 4, Application US/09636077A  
; Patent No. 6537785  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM  
; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE  
; FILE REFERENCE: 195612USO  
; CURRENT APPLICATION NUMBER: US/09/636,077A  
; CURRENT FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patent in version 3.0

; SEQ ID NO 4  
; LENGTH: 5597  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-636-077A-4

## Query Match

100.0%; Score 5597; DB 4; Length 5597;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGAGCCGAGCGGGCGTCCGTCGCGGAGCTGCAATGAGCGCGCCCGGAGCTGTGACC 60  
Db 1 CGAGCCGAGCGGGCGTCCGTCGCGGAGCTGCAATGAGCGCGCCCGGAGCTGTGACC 60  
Qy 61 TCGGCGCGCGCGCCGACCGGGGCCCTGAATGGCGGCTCGCTGAGCGCGCGCGCGGC 120  
Db 61 TCGGCGCGCGCGCCGACCGGGGCCCTGAATGGCGGCTCGCTGAGCGCGCGCGCGGC 120  
Qy 121 GCGCGCTCAGGCTCCTCGGGGCGTGGCGGTGAAGGGGTGATGCTGTTCAAGCTCC 180  
Db 121 GCGCGCTCAGGCTCCTCGGGGCGTGGCGGTGAAGGGGTGATGCTGTTCAAGCTCC 180  
Qy 181 TCGAGAGACAAACCTATACCTGCTGTCCACAGATATGGGCTCTACGTGCTTTTGG 240  
Db 181 TCGAGAGACAAACCTATACCTGCTGTCCACAGATATGGGCTCTACGTGCTTTTGG 240  
Qy 241 GCGTGTGTCACCATCTGCTCCGCTTCCAGTTCGGAGAGGTGTTCTGGAATGGAGCC 300  
Db 241 GCGTGTGTCACCATCTGCTCCGCTTCCAGTTCGGAGAGGTGTTCTGGAATGGAGCC 300  
Qy 301 GAGATCAATACCATGTTTGTGTTGATTCCTATAGAGACAAATATGCTGGAAGTCTTTC 360  
Db 301 GAGATCAATACCATGTTTGTGTTGATTCCTATAGAGACAAATATGCTGGAAGTCTTTC 360  
Qy 361 AGAATCGGCTTGTGTCGCCATCGGATTCGAGTGTGTTACACCTGGGTGATGGCAGC 420  
Db 361 AGAATCGGCTTGTGTCGCCATCGGATTCGAGTGTGTTACACCTGGGTGATGGCAGC 420  
Qy 421 ATCTTGAATCTACTGAAGAACTACAGAGCTCAGAGAACAGATGAGGAGGAGCAGAAAG 480  
Db 421 ATCTTGAATCTACTGAAGAACTACAGAGCTCAGAGAACAGATGAGGAGGAGCAGAAAG 480  
Qy 481 CAATGAGAGAAATCTTGGGAAAACACAAACGAACTTAAAGAGTGAAGCAGT 540  
Db 481 CAATGAGAGAAATCTTGGGAAAACACAAACGAACTTAAAGAGTGAAGCAGT 540  
Qy 541 TAGAGTGTGTTGTAACACACTGCAATTAAGGTGCCAATGCTTGTACTGGACCCGCTGC 600  
Db 541 TAGAGTGTGTTGTAACACACTGCAATTAAGGTGCCAATGCTTGTACTGGACCCGCTGC 600  
Qy 601 CAGCCAACTCAGCTGAGGAGCTGCCATCTTTTATCTTTTCCTTTTCCTTTTCAGTGT 660  
Db 601 CAGCCAACTCAGCTGAGGAGCTGCCATCTTTTATCTTTTCCTTTTCCTTTTCAGTGT 660  
Qy 661 ACATTTTCAATGTTGCAAAACCAAAACCCCTTACCAATGCTCAGTGTGTTTGG 720  
Db 661 ACATTTTCAATGTTGCAAAACCAAAACCCCTTACCAATGCTCAGTGTGTTTGG 720  
Qy 721 ACAGTACTAAGGATGTTGAAGATGCCACTCTGGAGTCTTTAAAGGAAATAGCAGAGA 780  
Db 721 ACAGTACTAAGGATGTTGAAGATGCCACTCTGGAGTCTTTAAAGGAAATAGCAGAGA 780  
Qy 781 CAGTATGAGGGGTACTTGAACAAGAGTCCCTGGATTTAGTGTATGCAAG 840  
Db 781 CAGTATGAGGGGTACTTGAACAAGAGTCCCTGGATTTAGTGTATGCAAG 840  
Qy 841 ATTTGGCTTTCTGAGTGGATTTCCACCACTTCAAGGAAACAAATCACTAAACAA 900  
Db 841 ATTTGGCTTTCTGAGTGGATTTCCACCACTTCAAGGAAACAAATCACTAAACAA 900  
Qy 901 AATTGCCAGAAATCTTTCTCTAAAGTCAAACTGTTGCAAGTGTATCAGAGCCAGT 960  
Db 901 AATTGCCAGAAATCTTTCTCTAAAGTCAAACTGTTGCAAGTGTATCAGAGCCAGT 960

QY 961 TAGCGCTTCTAAACTGAATTAACCCCAAGGATTTTCAAGATTTGAAATAAGCAAACTAAGA 1020  
Db 961 TAGCCCTTCTAAACTGAATTAACCCCAAGGATTTTCAAGATTTGAAATAAGCAAACTAAGA 1020  
QY 1021 AGAATGATGACCATGATGGAAGAAGAACTGACCATAGTCTGCGATATTTATTTATGGGATC 1080  
Db 1021 AGAATGATGACCATGATGGAAGAAGAACTGACCATAGTCTGCGATATTTATTTATGGGATC 1080  
QY 1081 TGAGCGCATGAGCCAGTCTAAGCAGGATGAAGACATCTGCGAGTGGTTTGAAGATA 1140  
Db 1081 TGAGCGCATGAGCCAGTCTAAGCAGGATGAAGACATCTGCGAGTGGTTTGAAGATA 1140  
QY 1141 ACGAAGAACTGAGGTACTCATTTGCGATCTATCGAGAGCATGACCATGGTTTGAAGATA 1200  
Db 1141 ACGAAGAACTGAGGTACTCATTTGCGATCTATCGAGAGCATGACCATGGTTTGAAGATA 1200  
QY 1201 TTTTCATTTGACCAACCGGAGATTTCCATCTGCGTGAACCTTGACAACTCTCGAGTGA 1260  
Db 1201 TTTTCATTTGACCAACCGGAGATTTCCATCTGCGTGAACCTTGACAACTCTCGAGTGA 1260  
QY 1261 CAATAGTAACACACAGGATGTTTTGCAAAATTTGAGGCACTTGCCCTACCTTTAGTTTCA 1320  
Db 1261 CAATAGTAACACACAGGATGTTTTGCAAAATTTGAGGCACTTGCCCTACCTTTAGTTTCA 1320  
QY 1321 CTGCTATTGAAAGTCAATTCATCGCATCGAGGCTGTCCCAAGGTTTATTTACCTAA 1380  
Db 1321 CTGCTATTGAAAGTCAATTCATCGCATCGAGGCTGTCCCAAGGTTTATTTACCTAA 1380  
QY 1381 ATGATGATGTCATGTTTGGAGAGGATGTCGSCAGATGATTTTACAGTCACCTCCAAG 1440  
Db 1381 ATGATGATGTCATGTTTGGAGAGGATGTCGSCAGATGATTTTACAGTCACCTCCAAG 1440  
QY 1441 GCCAGAAGGTTTATTTGATGCGCTGTGCCAACTGTCCGAGGCTGCCAGGTTCT 1500  
Db 1441 GCCAGAAGGTTTATTTGATGCGCTGTGCCAACTGTCCGAGGCTGCCAGGTTCT 1500  
QY 1501 GGAATTAAGATGGCTATTGTCACAGGCTTGTAAATTAATTCAGCTGCGATTTGGATGGTG 1560  
Db 1501 GGAATTAAGATGGCTATTGTCACAGGCTTGTAAATTAATTCAGCTGCGATTTGGATGGTG 1560  
QY 1561 GGGATTTGCTGGAACAGTGGAGGAGTGGCTATATTGCGAGGAGTGGAGTACTGGGA 1620  
Db 1561 GGGATTTGCTGGAACAGTGGAGGAGTGGCTATATTGCGAGGAGTGGAGTACTGGGA 1620  
QY 1621 GTATTGGAGTTGGACACCCCTGGCAGTTTGTGGAGGAATAAACAGTGTCTCTTACTGTA 1680  
Db 1621 GTATTGGAGTTGGACACCCCTGGCAGTTTGTGGAGGAATAAACAGTGTCTCTTACTGTA 1680  
QY 1681 ATCAGGATGTCGGAATTCCTGGCTCGCTGATAGTTCTGTGACCAAGCATGCAATGCT 1740  
Db 1681 ATCAGGATGTCGGAATTCCTGGCTCGCTGATAGTTCTGTGACCAAGCATGCAATGCT 1740  
QY 1741 TGTCTGTGGGTTTGATGTCGCGACTGTGGSCAAGATCATTTTTCATGAATTTGATAAAG 1800  
Db 1741 TGTCTGTGGGTTTGATGTCGCGACTGTGGSCAAGATCATTTTTCATGAATTTGATAAAG 1800  
QY 1801 TGATCCTTCTCCAAACAGATCTACTATATTATTTCBAAGGTGAATGCCCTGCTTAT 1860  
Db 1801 TGATCCTTCTCCAAACAGATCTACTATATTATTTCBAAGGTGAATGCCCTGCTTAT 1860  
QY 1861 TCAGCTTTTCAGAAAGTAGCCAAAGAGGAGTTGAAGGTGCTTATAGTGAACAATCCAAATA 1920  
Db 1861 TCAGCTTTTCAGAAAGTAGCCAAAGAGGAGTTGAAGGTGCTTATAGTGAACAATCCAAATA 1920  
QY 1921 TTGACATGCTTCTATTGCGAACAAGTGGAAACCATTCACCTCTAATATGACAGTGGAA 1980  
Db 1921 TTGACATGCTTCTATTGCGAACAAGTGGAAACCATTCACCTCTAATATGACAGTGGAA 1980  
QY 1981 TGAATGCCACCAATATACATTTTAACTCAAGTTTCAAAATCAAAACGATGAAGTTCA 2040  
Db 1981 TGAATGCCACCAATATACATTTTAACTCAAGTTTCAAAATCAAAACGATGAAGTTCA 2040

QY 2041 AAATGCAGATAACAGTGGAGTGGACAAGGAGGAGGACCAAACTGAATTTCTACGGCCC 2100  
Db 2041 AAATGCAGATAACAGTGGAGTGGACAAGGAGGAGGACCAAACTGAATTTCTACGGCCC 2100  
QY 2101 AGAAGGTTTACGAAATTTAGTTAGTCCCATTAACACTTCTTCCAGAGGCGGAATCTTT 2160  
Db 2101 AGAAGGTTTACGAAATTTAGTTAGTCCCATTAACACTTCTTCCAGAGGCGGAATCTTT 2160  
QY 2161 TTGAGATATTTCCAAAGAAACCGCTTCCGAAAGTTTAAAGAGACATGATGTTAACTCAA 2220  
Db 2161 TTGAGATATTTCCAAAGAAACCGCTTCCGAAAGTTTAAAGAGACATGATGTTAACTCAA 2220  
QY 2221 CAAGGAGAGCCCAAGGAGGTTGAAAATTTCCCTCGTAAATATTTTCACTCTTCCAAAAG 2280  
Db 2221 CAAGGAGAGCCCAAGGAGGTTGAAAATTTCCCTCGTAAATATTTTCACTCTTCCAAAAG 2280  
QY 2281 ACCCCAGTGTAGTCTCAATACCTGCAATTTGCAACTGGAACATGAGACATCCTTTTGA 2340  
Db 2281 ACCCCAGTGTAGTCTCAATACCTGCAATTTGCAACTGGAACATGAGACATCCTTTTGA 2340  
QY 2341 AAGGATACAAATTTGTCGAGTCTGAGATCATTTCTGATGAACTCAGACATG 2400  
Db 2341 AAGGATACAAATTTGTCGAGTCTGAGATCATTTCTGATGAACTCAGACATG 2400  
QY 2401 CTAATAATAAAATCAAGCTATTAACAGATGAACAAATGACAGTTTGGTGCTCCAC 2460  
Db 2401 CTAATAATAAAATCAAGCTATTAACAGATGAACAAATGACAGTTTGGTGCTCCAC 2460  
QY 2461 AGAAAAACAGGTTTCAAAAAGCATCTTCCCAACAGCTTAGGAGTGTCTGAAAGATTGC 2520  
Db 2461 AGAAAAACAGGTTTCAAAAAGCATCTTCCCAACAGCTTAGGAGTGTCTGAAAGATTGC 2520  
QY 2521 AGAGTTTGAATTTCTGAGTGTGAAAGTGAATGATGATGATGATGATGATGATGATGATG 2580  
Db 2521 AGAGTTTGAATTTCTGAGTGTGAAAGTGAATGATGATGATGATGATGATGATGATGATG 2580  
QY 2581 CACCTTGGATTTGGAGACCAAGCAAGATTTAGAGTGAACACTCACACCAAAACCA 2640  
Db 2581 CACCTTGGATTTGGAGACCAAGCAAGATTTAGAGTGAACACTCACACCAAAACCA 2640  
QY 2641 TAGCGGAAATGTGACAAAGAAAGCCCTCTCTGATTTGTTTCACTGGAAAGCCAGA 2700  
Db 2641 TAGCGGAAATGTGACAAAGAAAGCCCTCTCTGATTTGTTTCACTGGAAAGCCAGA 2700  
QY 2701 TGACAAAGAAAGAAATCACAGGAAAGAAAGAGAACAGTGAATGGAGGAAATG 2760  
Db 2701 TGACAAAGAAAGAAATCACAGGAAAGAAAGAGAACAGTGAATGGAGGAAATG 2760  
QY 2761 CTGAAATACATAGGCTTACTGAAAGTGTACTTGGAAAGAGCTGCAGATTACACAG 2820  
Db 2761 CTGAAATACATAGGCTTACTGAAAGTGTACTTGGAAAGAGCTGCAGATTACACAG 2820  
QY 2821 ATAGTTTACTTGGGCTTTTGGCATGGGAGAAAGAAAGTATTTTCCAGATCTTCTCGAGC 2880  
Db 2821 ATAGTTTACTTGGGCTTTTGGCATGGGAGAAAGAAAGTATTTTCCAGATCTTCTCGAGC 2880  
QY 2881 AAGAGAGTCAATTTGAAGACAAATTTGGCATTTCTGATAGCAAAATTTCTGGAGGC 2940  
Db 2881 AAGAGAGTCAATTTGAAGACAAATTTGGCATTTCTGATAGCAAAATTTCTGGAGGC 2940  
QY 2941 AACTAAAGATATATTTGAGATTTCCCTGATGATGATAAATAAATTTCTAAATAGCAAGT 3000  
Db 2941 AACTAAAGATATATTTGAGATTTCCCTGATGATGATAAATAAATTTCTAAATAGCAAGT 3000  
QY 3001 TTGATTTACATCGCGAAAGTCCCTGCTCACATGCTCACATGATTTGACCGGATTTGTTA 3060  
Db 3001 TTGATTTACATCGCGAAAGTCCCTGCTCACATGCTCACATGATTTGACCGGATTTGTTA 3060  
QY 3061 TGCAAGAACTGCAAGATATTTTCCCTGAAAGATTTGCAAGAGCTCATTTTCAAAAGTGC 3120  
Db 3061 TGCAAGAACTGCAAGATATTTTCCCTGAAAGATTTGCAAGAGCTCATTTTCAAAAGTGC 3120  
QY 3121 GCCATTCTGAGGATGACAGTTTGGCTTCTTATTTTATTTATTTCTCATGAGTGCAGTGC 3180



Db 3121 GCATCTCTGAGGATATGAGTTGGCTTCTCTTATTTTATTTATCTCATGAGTCAGTGC 3180  
Qy 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTCTCTGT 3240  
Db 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTCTCTGT 3240  
Qy 3241 CTGACAGAGAAATCCGAACACTGGCTACCGAAATTCAGAACTCCGTTAAAGTTTGAGG 3300  
Db 3241 CTGACAGAGAAATCCGAACACTGGCTACCGAAATTCAGAACTCCGTTAAAGTTTGAGG 3300  
Qy 3301 ATTTGACAGGTTCTGGAAACACATGCTTAATAATTCGTCAAAATGCTTCCTGCTGATATCA 3360  
Db 3301 ATTTGACAGGTTCTGGAAACACATGCTTAATAATTCGTCAAAATGCTTCCTGCTGATATCA 3360  
Qy 3361 GCGAGCTAAATTAATTCACCAACTCAGGAATCCTACTATGATCCCAACTGCCACCGG 3420  
Db 3361 GCGAGCTAAATTAATTCACCAACTCAGGAATCCTACTATGATCCCAACTGCCACCGG 3420  
Qy 3421 TCACCTAAGTCTAGTACCAACTGTAAACCAAGTAACTGACAAATCCACAAAGCATATA 3480  
Db 3421 TCACCTAAGTCTAGTACCAACTGTAAACCAAGTAACTGACAAATCCACAAAGCATATA 3480  
Qy 3481 AGGACAAAACAAATATAGGTTGAAATCATGGGAGAGAAATCGCTTTTAAATGA 3540  
Db 3481 AGGACAAAACAAATATAGGTTGAAATCATGGGAGAGAAATCGCTTTTAAATGA 3540  
Qy 3541 TTCTGACCAAGTTTCTCATGTTGGTGGCCAGTTGGATGACATGAAGAAAACCCTAGGA 3600  
Db 3541 TTCTGACCAAGTTTCTCATGTTGGTGGCCAGTTGGATGACATGAAGAAAACCCTAGGA 3600  
Qy 3601 AGTTGTTGCTGCTGAATGACAACTGACCAATCATAAAGATGCTCAGACAGTGAAG 3660  
Db 3601 AGTTGTTGCTGCTGAATGACAACTGACCAATCATAAAGATGCTCAGACAGTGAAG 3660  
Qy 3661 CTGTTCTCAGGAGCTTCTATGAATCCATGTTTCCCACTACCTTCCCAATTTGAACTGCCAA 3720  
Db 3661 CTGTTCTCAGGAGCTTCTATGAATCCATGTTTCCCACTACCTTCCCAATTTGAACTGCCAA 3720  
Qy 3721 CAGAGTATCGAAACCGTTTCTCATATGATGATGCTGCGAGTATGAGGCTTATCGAG 3780  
Db 3721 CAGAGTATCGAAACCGTTTCTCATATGATGATGCTGCGAGTATGAGGCTTATCGAG 3780  
Qy 3781 ACAAAATGAAGTTTGGACCCATTTGTACTAGCAACATTTGATTTTACTATATCT 3840  
Db 3781 ACAAAATGAAGTTTGGACCCATTTGTACTAGCAACATTTGATTTTACTATATCT 3840  
Qy 3841 CATTTTGTGCTGAGGATTAATTCGACTTAAGCGGAGATATTTCCAGAGAGGATAC 3900  
Db 3841 CATTTTGTGCTGAGGATTAATTCGACTTAAGCGGAGATATTTCCAGAGAGGATAC 3900  
Qy 3901 ACAAGAGCTAGTCCCAATCGAATCAGAGTATAGAAATCTTCATTTGAAACCATCTA 3960  
Db 3901 ACAAGAGCTAGTCCCAATCGAATCAGAGTATAGAAATCTTCATTTGAAACCATCTA 3960  
Qy 3961 CCTCAGCATTTACTGAGCATTTTAAATCTCAGCTTCAAGATGCTTTGTGATGTGAT 4020  
Db 3961 CCTCAGCATTTACTGAGCATTTTAAATCTCAGCTTCAAGATGCTTTGTGATGTGAT 4020  
Qy 4021 GCTTAGCAGTTTGGCCCGAAGAGAAATATCCAGTACCATGCTGTTTGTGGCATGAA 4080  
Db 4021 GCTTAGCAGTTTGGCCCGAAGAGAAATATCCAGTACCATGCTGTTTGTGGCATGAA 4080  
Qy 4081 TATAGCCCATGACTAGAAATTTTAAACCAACCCACTGAAAACTTGTGTGCGAGCAGC 4140  
Db 4081 TATAGCCCATGACTAGAAATTTTAAACCAACCCACTGAAAACTTGTGTGCGAGCAGC 4140  
Qy 4141 TCTGAATCTGATTTTACTTTTAAAGATTTGCTCATGACCTGCTGATCTTTTATAAAA 4200  
Db 4141 TCTGAATCTGATTTTACTTTTAAAGATTTGCTCATGACCTGCTGATCTTTTATAAAA 4200  
Qy 4201 GGCTCAGTCAAGAGACAGCTGTTAAATTTCCACAGCAATCATTTGACAGCTAACTTAT 4260

Db 4201 GGCTCAGTCAAGAGACAGCTGTTAAATTTCCACAGCAATCATTTGACAGCTAACTTAT 4260  
Qy 4261 TAGGAGAACCTTATGCGCAGCTGGAGTGAATCTCTAAGAGGCTCCAGTCTTTGCAATCCAA 4320  
Db 4261 TAGGAGAACCTTATGCGCAGCTGGAGTGAATCTCTAAGAGGCTCCAGTCTTTGCAATCCAA 4320  
Qy 4321 AGCCTTTGCTAAAGTTTGGCACTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 4380  
Db 4321 AGCCTTTGCTAAAGTTTGGCACTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 4380  
Qy 4381 AGTTAACTAGTTTATCTTCTCTCTGAGTATACGAATTTGGGATGCTTAAACCTATTTT 4440  
Db 4381 AGTTAACTAGTTTATCTTCTCTCTGAGTATACGAATTTGGGATGCTTAAACCTATTTT 4440  
Qy 4441 TAGATGTTTATTTAAATAATGACGAATATCACTCTTTATTTGACAAATCACTAAATTTATGAG 4500  
Db 4441 TAGATGTTTATTTAAATAATGACGAATATCACTCTTTATTTGACAAATCACTAAATTTATGAG 4500  
Qy 4501 TTTTATTAATATTTTAAAGTCTTAATGCTTTTAAACCACTACTACTCTGAGAGCTCAATG 4560  
Db 4501 TTTTATTAATATTTTAAAGTCTTAATGCTTTTAAACCACTACTACTCTGAGAGCTCAATG 4560  
Qy 4561 ATTGACATCTGAAATGCTTTTGTAAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGA 4620  
Db 4561 ATTGACATCTGAAATGCTTTTGTAAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGA 4620  
Qy 4621 CGTGCAGGCTTAATTTTCAACAGGCTAGAGTTAGTACTACTTACCAGATGTAATTTATGTTT 4680  
Db 4621 CGTGCAGGCTTAATTTTCAACAGGCTAGAGTTAGTACTACTTACCAGATGTAATTTATGTTT 4680  
Qy 4681 TCGAAATGTACATATTTCAACAGAGTGCCTCAATTTTGAATAATGAGTACTGATGCGCA 4740  
Db 4681 TCGAAATGTACATATTTCAACAGAGTGCCTCAATTTTGAATAATGAGTACTGATGCGCA 4740  
Qy 4741 CTGGCAGCATTAAGTGTGTCTTGTAAATTAATCTCATTTGGTATATTCAGTACTATCTCT 4800  
Db 4741 CTGGCAGCATTAAGTGTGTCTTGTAAATTAATCTCATTTGGTATATTCAGTACTATCTCT 4800  
Qy 4801 CTGAGTTGTTTTCATAGAACAGAGGCGCAGCAACTTTCTTTGTAAGGCTGGTGTAGT 4860  
Db 4801 CTGAGTTGTTTTCATAGAACAGAGGCGCAGCAACTTTCTTTGTAAGGCTGGTGTAGT 4860  
Qy 4861 AAATTTATTCAGGCGCACCTGTCTTGTCTATACATTTCTTGTGTGTGTGTGTGTGTGTGT 4920  
Db 4861 AAATTTATTCAGGCGCACCTGTCTTGTCTATACATTTCTTGTGTGTGTGTGTGTGTGTGT 4920  
Qy 4921 TTTTCTTCAACCAACCTCTTAAATGTAATAATGTAATAATGTAATAATGTAATAATGTAATAAT 4980  
Db 4921 TTTTCTTCAACCAACCTCTTAAATGTAATAATGTAATAATGTAATAATGTAATAATGTAATAAT 4980  
Qy 4981 ACTGCCACCCAGCAGATGTGACCCCTCAGGCGCATCTTTGCAATCACTGAGAAATTTT 5040  
Db 4981 ACTGCCACCCAGCAGATGTGACCCCTCAGGCGCATCTTTGCAATCACTGAGAAATTTT 5040  
Qy 5041 TTGTTTGT 5100  
Db 5041 TTGTTTGT 5100  
Qy 5101 TGCAGTGGCGCAATCTCAGCTCAGTCAACCTCTCGCCCTCCCGGGTTCAAGCAGTTCTGTCT 5160  
Db 5101 TGCAGTGGCGCAATCTCAGCTCAGTCAACCTCTCGCCCTCCCGGGTTCAAGCAGTTCTGTCT 5160  
Qy 5161 TCAGCCTTCTGAGTAGCTGGGACTACAGGTGATGCGCAACCACTCTGCTGCTAAATTTTGTGA 5220  
Db 5161 TCAGCCTTCTGAGTAGCTGGGACTACAGGTGATGCGCAACCACTCTGCTGCTAAATTTTGTGA 5220  
Qy 5221 TTTTGTAGTAGAGAGCGGGGTTTCCACCATATTTGGTCTGAGCTTATCTGAACTCTGACCTC 5280  
Db 5221 TTTTGTAGTAGAGAGCGGGGTTTCCACCATATTTGGTCTGAGCTTATCTGAACTCTGACCTC 5280  
Qy 5281 AGGTGATCCACTGCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 5340  
Db 5281 AGGTGATCCACTGCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 5340



QY 5341 CCAGCCGAGAAATTAGTATTTTATGTTATGTTAAACCTTGGCTCTAGCCATATTTTATG 5400  
DB 5341 CCAGCCGAGAAATTAGTATTTTATGTTATGTTAAACCTTGGCTCTAGCCATATTTTATG 5400  
QY 5401 TCATAAATCAATGGATTTGGAAGAGAGATTCATAGTAACTCTGACAGGATTTTATG 5460  
DB 5401 TCATAAATCAATGGATTTGGAAGAGAGATTCATAGTAACTCTGACAGGATTTTATG 5460  
QY 5461 ATCATGATCTCAACAATATTCCTCCAAATGGCATACATCTTTTGTACAAAGAACTTGAA 5520  
DB 5461 ATCATGATCTCAACAATATTCCTCCAAATGGCATACATCTTTTGTACAAAGAACTTGAA 5520  
QY 5521 ATGTAATAATCTGTTGTTGCTGTGAAGATTTGTGTTATTTCAAAACTGAAATCTCATAAA 5580  
DB 5521 ATGTAATAATCTGTTGTTGCTGTGAAGATTTGTGTTATTTCAAAACTGAAATCTCATAAA 5580  
QY 5581 AAGTTAAATTTTGA AAA 5597  
DB 5581 AAGTTAAATTTTGA AAA 5597

RESULT 3  
US-09-636-060C-4  
; Sequence 4, Application US/09636060C  
; Patent No. 6642038  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM M  
; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY  
; FILE REFERENCE: 210119USOCONT  
; CURRENT APPLICATION NUMBER: US/09/636,060C  
; CURRENT FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 5597  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-636-060C-4

Query Match 100.0%; Score 5597; DB 4; Length 5597;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CGGAGCCGAGCGGGCGTCCGTGCGCGAGCTGCAATGAGCGCGCGCGAGCGCTGTGACC 60  
DB 1 CGGAGCCGAGCGGGCGTCCGTGCGCGAGCTGCAATGAGCGCGCGCGAGCGCTGTGACC 60  
QY 61 TCG 120  
DB 61 TCG 120  
QY 121 GCGCGCTCAGGCTCCTCGCGCGCGTGGCGTGGCGGTGAAGGGGTGATGCTGTTCAAGTCC 180  
DB 121 GCGCGCTCAGGCTCCTCGCGCGCGTGGCGTGGCGGTGAAGGGGTGATGCTGTTCAAGTCC 180  
QY 181 TCGAGAGCAAACTATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
DB 181 TCGAGAGCAAACTATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
QY 241 GCGTGGTGTACCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300  
DB 241 GCGTGGTGTACCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300  
QY 301 GAGATCAATACATGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 360  
DB 301 GAGATCAATACATGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 360  
QY 361 AGAATCGGCTTGTCTGCG 420  
DB 361 AGAATCGGCTTGTCTGCG 420

QY 421 ATCTTGAACTACTGAAGGAACCTACAGCAGGTACAGAAACAGATGAGGAGGAGCAGAAAG 480  
DB 421 ATCTTGAACTACTGAAGGAACCTACAGCAGGTACAGAAACAGATGAGGAGGAGCAGAAAG 480  
QY 481 CAATGAGAGAAATCCCTTGGGAAAAACACAAACGGAACCTACTAAGAGAGATGAGAGCAGT 540  
DB 481 CAATGAGAGAAATCCCTTGGGAAAAACACAAACGGAACCTACTAAGAGAGATGAGAGCAGT 540  
QY 541 TAGAGTGTGTTGCTTAACACACTGATTAAGGTGCCAATGCTTGTACTTGGACCCAGCCCTGC 600  
DB 541 TAGAGTGTGTTGCTTAACACACTGATTAAGGTGCCAATGCTTGTACTTGGACCCAGCCCTGC 600  
QY 601 CAGCCAACTACACCTTGAAGGAGGTGCCATCTCTTTATCTTTTCTTTTCTTTTCTTTTCT 660  
DB 601 CAGCCAACTACACCTTGAAGGAGGTGCCATCTCTTTATCTTTTCTTTTCTTTTCTTTTCT 660  
QY 661 ACATTTTCAATGTTGCAAAAACCAAAAACCCCTTCTACCAATGCTCTCAGTGTGTTGTTT 720  
DB 661 ACATTTTCAATGTTGCAAAAACCAAAAACCCCTTCTACCAATGCTCTCAGTGTGTTGTTT 720  
QY 721 ACAGTACTAAGGATGTTGAAGATGCCCACTCTGGACTGCTTAAAGGAAATAGCAGACAGA 780  
DB 721 ACAGTACTAAGGATGTTGAAGATGCCCACTCTGGACTGCTTAAAGGAAATAGCAGACAGA 780  
QY 781 CAGTATGAGGGGGTACTTGAACAGATAAAGAGTCCCTGGATTTAGTCTTAATGCAAG 840  
DB 781 CAGTATGAGGGGGTACTTGAACAGATAAAGAGTCCCTGGATTTAGTCTTAATGCAAG 840  
QY 841 ATTTGCTTTCTCTGAGTGATTTCCACCACATTTCAAGGAAACAAATCAATCAATCAAA 900  
DB 841 ATTTGCTTTCTCTGAGTGATTTCCACCACATTTCAAGGAAACAAATCAATCAATCAAA 900  
QY 901 AATTGCGAGAAATCTTCTCTAAAGTCAAACTGTTGAGTGTGATTTAGAGGCGCAGTG 960  
DB 901 AATTGCGAGAAATCTTCTCTAAAGTCAAACTGTTGAGTGTGATTTAGAGGCGCAGTG 960  
QY 961 TAGCGCTTCTAATACTGAATTAACCCCAAGGATTTTCAAGAATTTGAATAGCAAACTAAGA 1020  
DB 961 TAGCGCTTCTAATACTGAATTAACCCCAAGGATTTTCAAGAATTTGAATAGCAAACTAAGA 1020  
QY 1021 AGAATCATGACATTTGATGGAAGAAGAACTGACCAATAGTCTCTGCATATTTATTATGGATC 1080  
DB 1021 AGAATCATGACATTTGATGGAAGAAGAACTGACCAATAGTCTCTGCATATTTATTATGGATC 1080  
QY 1081 TGAGCGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCGAGTCTGTTTGAAGATA 1140  
DB 1081 TGAGCGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCGAGTCTGTTTGAAGATA 1140  
QY 1141 ACCAAGAACTGAGGTACTCAATTGCGATCTATCGAGAGGATGCAACATGGGTTCCGAAATA 1200  
DB 1141 ACCAAGAACTGAGGTACTCAATTGCGATCTATCGAGAGGATGCAACATGGGTTCCGAAATA 1200  
QY 1201 TTTTTCATTTGTCACCAACGGGCGAGATTCATCTCTGGTGAACCTTGAACAATCTCTCGAGTGA 1260  
DB 1201 TTTTTCATTTGTCACCAACGGGCGAGATTCATCTCTGGTGAACCTTGAACAATCTCTCGAGTGA 1260  
QY 1261 CATAGTAAACACACAGGATGTTTTCGAAATTTGAGCCACTTGCCTACCTTTAGTTTAC 1320  
DB 1261 CATAGTAAACACACAGGATGTTTTCGAAATTTGAGCCACTTGCCTACCTTTAGTTTAC 1320  
QY 1321 CTGCTATTGAAAGTCAATTCATCGCATCGAAGGGCTGTCCAGAGAGTTTATTATTCCTAA 1380  
DB 1321 CTGCTATTGAAAGTCAATTCATCGCATCGAAGGGCTGTCCAGAGAGTTTATTATTCCTAA 1380  
QY 1381 ATGATGATGATGTTTGGAGGATGTTGCGCAGATGATTTTACAGTCACTTCCAAAG 1440  
DB 1381 ATGATGATGATGTTTGGAGGATGTTGCGCAGATGATTTTACAGTCACTTCCAAAG 1440  
QY 1441 GCCAGAGAGTTTATTTGACATGGCTGTGCCAAACTGTGCCAGGCGCTCCCGAGGTTCT 1500  
DB 1441 GCCAGAGAGTTTATTTGACATGGCTGTGCCAAACTGTGCCAGGCGCTCCCGAGGTTCT 1500

QY	1501	GGATTAAGGATGGCTATTGTGCAAGGCTTGTAATAATTCAGCTGCGATTTGGGATGGTG	1560
Db	1501	GGATTAAGGATGGCTATTGTGCAAGGCTTGTAATAATTCAGCTGCGATTTGGGATGGTG	1560
QY	1561	GGGATTGCTCTGGAAACAGTGGAGGAGTCCCTATATTGCGAGGAGTGGAGGTACTGGGA	1520
Db	1561	GGGATTGCTCTGGAAACAGTGGAGGAGTCCCTATATTGCGAGGAGTGGAGGTACTGGGA	1520
QY	1621	GTATTGGAGTTGGACACCCCTGGCAGTTTGGTGAGGAAATAAACAGTCTCTTACTGTGA	1680
Db	1621	GTATTGGAGTTGGACACCCCTGGCAGTTTGGTGAGGAAATAAACAGTCTCTTACTGTGA	1680
QY	1681	ATCAGGATGTCGCAATTCCTGGCTCGCTGATATAGTTCTGTGACCAAGCATGCAATGTCT	1740
Db	1681	ATCAGGATGTCGCAATTCCTGGCTCGCTGATATAGTTCTGTGACCAAGCATGCAATGTCT	1740
QY	1741	TGTCCTGTGGGTTTGATCTGGCACTGTGGCAAGATCATTTTCATGAATTTGTATAAG	1800
Db	1741	TGTCCTGTGGGTTTGATCTGGCACTGTGGCAAGATCATTTTCATGAATTTGTATAAG	1800
QY	1801	TGATCCTTCTCCAAACAGACTCACTATATATTCCAAAGAGTGAATGCTGCTCTTATT	1860
Db	1801	TGATCCTTCTCCAAACAGACTCACTATATATTCCAAAGAGTGAATGCTGCTCTTATT	1860
QY	1861	TCAGCTTTCAGAGTAGCCAAAGAGAGTGGAGTGGCTATAGTGACAATCCCAATAA	1920
Db	1861	TCAGCTTTCAGAGTAGCCAAAGAGAGTGGAGTGGCTATAGTGACAATCCCAATAA	1920
QY	1921	TTCCGATGCTTCTATTGCCAACAGTGGAAACCATCCCTCATATATGACACAGTGGAA	1980
Db	1921	TTCCGATGCTTCTATTGCCAACAGTGGAAACCATCCCTCATATATGACACAGTGGAA	1980
QY	1981	TGAATGCCACACATATATTTAATCTCAGTTTCAAAATACAAACGATGAGAGTTCA	2040
Db	1981	TGAATGCCACACATATATTTAATCTCAGTTTCAAAATACAAACGATGAGAGTTCA	2040
QY	2041	AAATGCAGATAACAGTGGAGTGGACACAAAGGAGGACCAAACTGAATTTCTACGGCCC	2100
Db	2041	AAATGCAGATAACAGTGGAGTGGACACAAAGGAGGACCAAACTGAATTTCTACGGCCC	2100
QY	2101	AGAGGGTTAGCAAAATTTAGTTAGTCCATAACACTTTCTCCAGAGGCGGAAATCCTTT	2160
Db	2101	AGAGGGTTAGCAAAATTTAGTTAGTCCATAACACTTTCTCCAGAGGCGGAAATCCTTT	2160
QY	2161	TTGAGGATTTCCCAAGAAAAACGTTCCCGAAGTTTAAGACATGATGTTAACTCAA	2220
Db	2161	TTGAGGATTTCCCAAGAAAAACGTTCCCGAAGTTTAAGACATGATGTTAACTCAA	2220
QY	2221	CAAGGAGGCCAGGAGTGAATAATTCCTCGTGGTAAATATTCTACTCCTTCCAAAAG	2280
Db	2221	CAAGGAGGCCAGGAGTGAATAATTCCTCGTGGTAAATATTCTACTCCTTCCAAAAG	2280
QY	2281	ACGCCAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAACATGAGACATCACTTTGA	2340
Db	2281	ACGCCAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAACATGAGACATCACTTTGA	2340
QY	2341	AAGGATCAAAATTTGTCCAAGTTCAGCTTGTGATGATCATTTCTGATGAATTCACAGATG	2400
Db	2341	AAGGATCAAAATTTGTCCAAGTTCAGCTTGTGATGATCATTTCTGATGAATTCACAGATG	2400
QY	2401	CTAAAAATAAAATCAAGCTATATAACAGATGAACAAATGACAGTTTGGTGGCTCCAC	2460
Db	2401	CTAAAAATAAAATCAAGCTATATAACAGATGAACAAATGACAGTTTGGTGGCTCCAC	2460
QY	2461	AGGAAAAACAGGTTCAATAAAGCATCTTGGCCAAACAGCTTAGGAGTGTCTGAAAGATTGC	2520
Db	2461	AGGAAAAACAGGTTCAATAAAGCATCTTGGCCAAACAGCTTAGGAGTGTCTGAAAGATTGC	2520
QY	2521	AGAGGTTGACTTTTCTCCAGTGTGATTAAGTGAATGTCATGACCCAGGCTCAGAAATC	2580
Db	2521	AGAGGTTGACTTTTCTCCAGTGTGATTAAGTGAATGTCATGACCCAGGCTCAGAAATC	2580
QY	2581	CACCCCTGGACTTTGGAGACCACAGCAAGATTTAGAGTGGAAACTCACACCCAAAAACCA	2640

Db	2581	CACCCCTGGACTTTGGAGACCACAGCAAGATTTAGAGTGGAAACTCACACCCAAAAACCA	2640
QY	2641	TAGGCGGAAATCTGACAAAAGAAAAGCCCCCATCTCTGATTTGTTCCACTGCAAAAGCCAGA	2700
Db	2641	TAGGCGGAAATCTGACAAAAGAAAAGCCCCCATCTCTGATTTGTTCCACTGCAAAAGCCAGA	2700
QY	2701	TCACAAAAGAAAAGAAAATTCACAGGAAAAGAAAAGAAACAGTAGAATGAGGAAAATG	2760
Db	2701	TCACAAAAGAAAAGAAAATTCACAGGAAAAGAAAAGAAACAGTAGAATGAGGAAAATG	2760
QY	2761	CTGAAAATCACATAGCGCTTACTGAAGTGTACTTTGGAGAAAGCTGAGCATTACACAG	2820
Db	2761	CTGAAAATCACATAGCGCTTACTGAAGTGTACTTTGGAGAAAGCTGAGCATTACACAG	2820
QY	2821	ATAGTTACTTTGGGCTTTTGGCCATGGGAGAAAAGAAAAGTATTTCCAGATCTTCTCGACG	2880
Db	2821	ATAGTTACTTTGGGCTTTTGGCCATGGGAGAAAAGAAAAGTATTTCCAGATCTTCTCGACG	2880
QY	2881	AAGAGAGTCATTGAGACACAAATTTGGCATCTTCACTGATAGCAAAAATCTGGGAGGC	2940
Db	2881	AAGAGAGTCATTGAGACACAAATTTGGCATCTTCACTGATAGCAAAAATCTGGGAGGC	2940
QY	2941	AACTAAAAGATACATTTGCGAGATTCCTCAGATATGTAATAAATAAATTTCTAATAGCAAGT	3000
Db	2941	AACTAAAAGATACATTTGCGAGATTCCTCAGATATGTAATAAATAAATTTCTAATAGCAAGT	3000
QY	3001	TTGGATTCACATTCGGGAAAAGTCCCTGCTCAGTCATGCCTCAATGATGACCGGATTTGTA	3060
Db	3001	TTGGATTCACATTCGGGAAAAGTCCCTGCTCAGTCATGCCTCAATGATGACCGGATTTGTA	3060
QY	3061	TCGAAAGATCTGCAAGATATTTGCTCCTGAAAGAAATTTGCAAGACGTCATTTCAAAAAGTGC	3120
Db	3061	TCGAAAGATCTGCAAGATATTTGCTCCTGAAAGAAATTTGCAAGACGTCATTTCAAAAAGTGC	3120
QY	3121	GCCATTTGAGGATATGAGTTTGGCTTCTCTTATTTTATTTATCTCAAGTGGAGTGC	3180
Db	3121	GCCATTTGAGGATATGAGTTTGGCTTCTCTTATTTTATTTATCTCAAGTGGAGTGC	3180
QY	3181	AGCCACTGAATATATCTCAAGTCTTTGATGAGTTGATACAGATCAATCTCTGTCTTGT	3240
Db	3181	AGCCACTGAATATATCTCAAGTCTTTGATGAGTTGATACAGATCAATCTCTGTCTTGT	3240
QY	3241	CTGACAGAGAAATCCGAAACACCTGCTACAGAAATTCAGAACTCCGTTAAATTTGACGG	3300
Db	3241	CTGACAGAGAAATCCGAAACACCTGCTACAGAAATTCAGAACTCCGTTAAATTTGACGG	3300
QY	3301	ATTGACAGGTTGAAACACATCTTAATAAATGCTGAAAATGCTTCTCTGATATCA	3360
Db	3301	ATTGACAGGTTGAAACACATCTTAATAAATGCTGAAAATGCTTCTCTGATATCA	3360
QY	3361	CCGAGTAAATAATATTCACCAACTCAGGAATCTACTATGATCCCAACCTGCCACCGG	3420
Db	3361	CCGAGTAAATAATATTCACCAACTCAGGAATCTACTATGATCCCAACCTGCCACCGG	3420
QY	3421	TCCTAAAAGTCTAGTAAACAACTGTAACCCAGTAACTGCAAAAATCCCAAGCATATA	3480
Db	3421	TCCTAAAAGTCTAGTAAACAACTGTAACCCAGTAACTGCAAAAATCCCAAGCATATA	3480
QY	3481	AGGACAAAACAAATATAGTTTGAATATCTGGGAGAAAGAAATCCCTTTTAAATGA	3540
Db	3481	AGGACAAAACAAATATAGTTTGAATATCTGGGAGAAAGAAATCCCTTTTAAATGA	3540
QY	3541	TTTCGTACCAAGTCTTCTCATGTGGTGGCCAGTTGGATGACATAAGAAAAACCCCTAGGA	3600
Db	3541	TTTCGTACCAAGTCTTCTCATGTGGTGGCCAGTTGGATGACATAAGAAAAACCCCTAGGA	3600
QY	3601	AGTTTGTGCTGAAATGACACATTTGACCAATCATAAAGATGCTCAGACAGTGAAGG	3660
Db	3601	AGTTTGTGCTGAAATGACACATTTGACCAATCATAAAGATGCTCAGACAGTGAAGG	3660
QY	3661	CTGTTCTCAGGAGCTTCTATGAATTCATGTTTCCCATACCTTCCCAATTTGAATGCGCAA	3720
Db	3661	CTGTTCTCAGGAGCTTCTATGAATTCATGTTTCCCATACCTTCCCAATTTGAATGCGCAA	3720

3661	CTGTTCTCAGGACCTCTCATGAATCCATGTTTCCCATACCTCTCCCAATTGAACTGCGAA	3721
3721	GAGAGTATCGAAACCGTTTCCTTCATATGCATGAGCTGCAGGAATGAGGGCTTATCGAG	3780
3721	GAGAGTATCGAAACCGTTTCCTTCATATGCATGAGCTGCAGGAATGAGGGCTTATCGAG	3780
3781	ACAAATCGAGTTTTTGACCCCATGCTGCTACTAGCAACAATTGATTTATGTTTACTATATCT	3840
3781	ACAAATCGAGTTTTTGACCCCATGCTGCTACTAGCAACAATTGATTTATGTTTACTATATCT	3840
3841	CATTTTTTGCTGAGCAGTTAATTCGACTTAAGCGGAAGATATTTCCGAGAAGGAGGATAC	3900
3841	CATTTTTTGCTGAGCAGTTAATTCGACTTAAGCGGAAGATATTTCCGAGAAGGAGGATAC	3900
3901	ACAAAGAAGCTAGTCCCAATCGAATCAGAGTATAGAAGATCTTCATTTGAAACCATCTA	3960
3901	ACAAAGAAGCTAGTCCCAATCGAATCAGAGTATAGAAGATCTTCATTTGAAACCATCTA	3960
3961	CCTCAGCAATTACTGAGCATTTTAAAACTCAGCTTTCACAGAGATGCTTTTGTGATGTGAT	4020
3961	CCTCAGCAATTACTGAGCATTTTAAAACTCAGCTTTCACAGAGATGCTTTTGTGATGTGAT	4020
4021	GCTTAGCAGTTTGGCCCGAAGAGAAAAATATCCAGTACCATGCTGTTTTGTGGCATGAA	4080
4021	GCTTAGCAGTTTGGCCCGAAGAGAAAAATATCCAGTACCATGCTGTTTTGTGGCATGAA	4080
4081	TATAGCCCACCTGACTAGGAATTAATTAACCAACCCACTGAAAACTTGCTGTCGAGCAGC	4140
4081	TATAGCCCACCTGACTAGGAATTAATTAACCAACCCACTGAAAACTTGCTGTCGAGCAGC	4140
4141	TCTGAACCTGATTTTACTTTTTAAAAAATTTGCTCATGACCTGTCATCCTTTTTTATAAAA	4200
4141	TCTGAACCTGATTTTACTTTTTAAAAAATTTGCTCATGACCTGTCATCCTTTTTTATAAAA	4200
4201	GGCTCAGTCAGACAGACAGCTGTTAATTTCCACAGCAATCATTCGAGACTAACTTTAT	4260
4201	GGCTCAGTCAGACAGACAGCTGTTAATTTCCACAGCAATCATTCGAGACTAACTTTAT	4260
4261	TAGGAGAAGCCTATGCCAGCTGGGAGTGATGCTTAAGAGAGCTCCAGTCTTGCATCCAA	4320
4261	TAGGAGAAGCCTATGCCAGCTGGGAGTGATGCTTAAGAGAGCTCCAGTCTTGCATCCAA	4320
4321	AGCCTTTTGCTAAAGTTTGGACTTTTTTTTTTTTTCATTTCCGATTTTAAAGTAGTAGTACTA	4380
4321	AGCCTTTTGCTAAAGTTTGGACTTTTTTTTTTTTTCATTTCCGATTTTAAAGTAGTAGTACTA	4380
4381	AGTTAACTAGTTATTTCTTGCTTCGAGTATAACGAAATGGGATGCTTAAACCTATTATTTA	4440
4381	AGTTAACTAGTTATTTCTTGCTTCGAGTATAACGAAATGGGATGCTTAAACCTATTATTTA	4440
4441	TAGATGTTATTTAAATAATGCGAATAATCACCTCTTATTGCAAAATACCTTAAATATGAG	4500
4441	TAGATGTTATTTAAATAATGCGAATAATCACCTCTTATTGCAAAATACCTTAAATATGAG	4500
4501	TTTTTATTAATTTTAAGACTGTAATCGTCTTAAACCACTAACTACTGGAAGAGCTCAATG	4560
4501	TTTTTATTAATTTTAAGACTGTAATCGTCTTAAACCACTAACTACTGGAAGAGCTCAATG	4560
4561	ATTGACATCTGAAATGCTTTGTAATTAATGACTTTCAGCCCCCTAAGAAATGCTATGATTTCA	4620
4561	ATTGACATCTGAAATGCTTTGTAATTAATGACTTTCAGCCCCCTAAGAAATGCTATGATTTCA	4620
4621	CGTGCAGGTCTAATTTCAAAGGCTAGAGTTAGTACTACTTACCCAGATGCTAATATGTTTT	4680
4621	CGTGCAGGTCTAATTTCAAAGGCTAGAGTTAGTACTACTTACCCAGATGCTAATATGTTTT	4680
4681	TGGAAATGCTACATATTTCAACAGAAAGTGCCTCATTTTTAGAAAATGAGTAGTGCATGGCA	4740
4681	TGGAAATGCTACATATTTCAACAGAAAGTGCCTCATTTTTAGAAAATGAGTAGTGCATGGCA	4740
4741	CTGGCAATTTACAGTGTGCTGTTTAAATATCTCATTTGGTATATTTCCAGTAGCTATCTCT	4800
4741	CTGGCAATTTACAGTGTGCTGTTTAAATATCTCATTTGGTATATTTCCAGTAGCTATCTCT	4800

## RESULT 4

RESULT 4  
US-09-986-552-4

US-09-386-332-4  
: Sequence 4. Application US/09986552

: Patent No. 6670165

FACEID NO. 0070103  
; GENERAL INFORMATION:

: : APPLICANT: CANFIELD, William

TITLE OF INVENTION: METHODS FOR PRODUCING

FILE REFERENCE: 215089US77DIV

CURRENT APPLICATION NUMBER: US/09/986,552

CURRENT FILING DATE: 2001-1-

; PRIOR APPLICATION NUMBER: 09/635,872

; PRIOR FILING DATE: 2000-08-10

; PRIORITY APPLICATION NUMBER: 60/153,831

; PRIOR FILING DATE: 1999-09-14

; NUMBER OF SEQ ID NOS: 52

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 5597
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-986-552-4

Query Match      100.0%; Score 5597; DB 4; Length 5597;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5597; Conserved 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGAGCCGAGGGGCGCTCCGTCGGCGAGCTGCAATGAGCGCGCCCGGAGGCTGTGACC 60
DB 1 CCGAGCCGAGGGGCGCTCCGTCGGCGAGCTGCAATGAGCGCGCCCGGAGGCTGTGACC 60
QY 61 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
DB 61 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
QY 121 GCGCGCTCAGGCTCTCTCGGCGCGCTGCGCGCGCGCTGCGCGCGCGCTGCGCGCGCG 180
DB 121 GCGCGCTCAGGCTCTCTCGGCGCGCTGCGCGCGCGCTGCGCGCGCGCTGCGCGCGCG 180
QY 181 TGCAGAGACAAACCTATACCTGCGCTGCCGCTCCACAGGATGCGCTACGCTGCTTTGG 240
DB 181 TGCAGAGACAAACCTATACCTGCGCTGCCGCTCCACAGGATGCGCTACGCTGCTTTGG 240
QY 241 GCGTGGTGTGACCATCGCTCCGCGCTCCAGTTCGAGAGGCTGCTTGGATGAGGCC 300
DB 241 GCGTGGTGTGACCATCGCTCCGCGCTCCAGTTCGAGAGGCTGCTTGGATGAGGCC 300
QY 301 GAGATCAATACCATGTTTGTGTTGATTCCTATAGAGACAAATATGCTGGAAAGTCTTTTC 360
DB 301 GAGATCAATACCATGTTTGTGTTGATTCCTATAGAGACAAATATGCTGGAAAGTCTTTTC 360
QY 361 AGAATCGGCTTTGTGCGCCATGCGGATGACGTTGTTTACACCTGGGTGAATGGGACAG 420
DB 361 AGAATCGGCTTTGTGCGCCATGCGGATGACGTTGTTTACACCTGGGTGAATGGGACAG 420
QY 421 ATCTTGAATCTACTGAAGAACTACAGCAGGTCAGAGAAACAGATGAGGAGGAGCAGAAAG 480
DB 421 ATCTTGAATCTACTGAAGAACTACAGCAGGTCAGAGAAACAGATGAGGAGGAGCAGAAAG 480
QY 481 CAATGAGAGAAATCTTTGGGAAAACACAAACGGAACCTCTAAGAAAGTGAAGGAGT 540
DB 481 CAATGAGAGAAATCTTTGGGAAAACACAAACGGAACCTCTAAGAAAGTGAAGGAGT 540
QY 541 TAGAGTGTGCTTAACACACTGCATTAGGTCGCAATGCTTGTACTGCAACCCAGCCCTGC 600
DB 541 TAGAGTGTGCTTAACACACTGCATTAGGTCGCAATGCTTGTACTGCAACCCAGCCCTGC 600
QY 601 CAGCCAAACATCACCTGGAAGGACGTCATCTCTTTATCTCTTTTCAATCTGCGCAGTG 660
DB 601 CAGCCAAACATCACCTGGAAGGACGTCATCTCTTTATCTCTTTTCAATCTGCGCAGTG 660
QY 661 ACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCAGTGTGTTGTTTGG 720
DB 661 ACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCAGTGTGTTGTTTGG 720
QY 721 ACAGTACTAAGGATGTTGAAGATGCCACTCTGAGCTGCTTAAAGGAAATAGCAGACAGA 780
DB 721 ACAGTACTAAGGATGTTGAAGATGCCACTCTGAGCTGCTTAAAGGAAATAGCAGACAGA 780
QY 781 CAGTATGAGGGGGTACTTGTGACACAGATAAAGAGTCCCTGGATTAAGTCTTAATGCAAG 840
DB 781 CAGTATGAGGGGGTACTTGTGACACAGATAAAGAGTCCCTGGATTAAGTCTTAATGCAAG 840
QY 841 ATTTGGCTTTCTCAGTGGATTTCCACCAACATTCAGAGGAAACAAATCAACTAAACAA 900
DB 841 ATTTGGCTTTCTCAGTGGATTTCCACCAACATTCAGAGGAAACAAATCAACTAAACAA 900
QY 901 AATTGCCAGAAATCTTTTCCTCTAAAGTCAAACCTGTTGAGTGTGATTCAGAGCCAGTG 960
DB 901 AATTGCCAGAAATCTTTTCCTCTAAAGTCAAACCTGTTGAGTGTGATTCAGAGCCAGTG 960

901 AATTGCCAGAAATCTTTTCCTCTAAAGTCAAACCTGTTGAGTGTGATTCAGAGCCAGTG 960
961 TAGCGCTTCTAAACTGAATTAACCCCAAGGATTTTCAAGAAATTAAGCAAACTAAGA 1020
961 TAGCGCTTCTAAACTGAATTAACCCCAAGGATTTTCAAGAAATTAAGCAAACTAAGA 1020
1021 AGAATGATGACCATGATGAGAAAGAACTGACATAGTCTCTGATATTTATTTATGGGATC 1080
1021 AGAATGATGACCATGATGAGAAAGAACTGACATAGTCTCTGATATTTATTTATGGGATC 1080
1081 TGAGGCCCATCAGCCAGTCTAAGCAGAGTGAAGACATCTCTCCAGTCTGTTGAAGATA 1140
1081 TGAGGCCCATCAGCCAGTCTAAGCAGAGTGAAGACATCTCTCCAGTCTGTTGAAGATA 1140
1141 ACGAAGAACTGAGGTACTCATTTGCGATCTATCGAGAGCAATGACCATGGGTTTGGAAATA 1200
1141 ACGAAGAACTGAGGTACTCATTTGCGATCTATCGAGAGCAATGACCATGGGTTTGGAAATA 1200
1201 TTTTCATTTGTCACCAACCGGCGAGATTCATCTCTGCTGAACTTTGACAACTCTCAGTGA 1260
1201 TTTTCATTTGTCACCAACCGGCGAGATTCATCTCTGCTGAACTTTGACAACTCTCAGTGA 1260
1261 CAATAGTAAACACACAGGATGTTTTCGAAATTTGAGCACTTGCCTACCTTTAGTTTAC 1320
1261 CAATAGTAAACACACAGGATGTTTTCGAAATTTGAGCACTTGCCTACCTTTAGTTTAC 1320
1321 CTGCTATTGAAAGTCACATTCATCGCATCGAAGGCTGTCGAGAGGATTTATTTACCTTAA 1380
1321 CTGCTATTGAAAGTCACATTCATCGCATCGAAGGCTGTCGAGAGGATTTATTTACCTTAA 1380
1381 ATGATGATGTCATGTTTGGGAAAGATGTCGCGCAGATGATTTTACAGTCACTCCAAAG 1440
1381 ATGATGATGTCATGTTTGGGAAAGATGTCGCGCAGATGATTTTACAGTCACTCCAAAG 1440
1441 GCCAGAGGTTTATTTGATGCGCTGTCGCAACTGTCGCGAGGCTGCCAGGCTTCTTCT 1500
1441 GCCAGAGGTTTATTTGATGCGCTGTCGCAACTGTCGCGAGGCTGCCAGGCTTCTTCT 1500
1501 GGATTAAGGATGGCTTATTTGCAAGGCTTGTAAATTAATTCAGCTTGGATGGATGGTG 1560
1501 GGATTAAGGATGGCTTATTTGCAAGGCTTGTAAATTAATTCAGCTTGGATGGATGGTG 1560
1561 GGGATGCTCTGGAAACAGTGGAGGAGTTCGCTATTTGCGAGGAGTGGAGGATCTGGGA 1620
1561 GGGATGCTCTGGAAACAGTGGAGGAGTTCGCTATTTGCGAGGAGTGGAGGATCTGGGA 1620
1621 GTATTGGAGTGGACACCCCTGGCAGTTTGGTGGAGGATTAACAGTGTCTCTTACTGTA 1680
1621 GTATTGGAGTGGACACCCCTGGCAGTTTGGTGGAGGATTAACAGTGTCTCTTACTGTA 1680
1681 ATCAGGATGTCGGAATTTCTGGCTGCTGATTAAGTTCGTGACCAAGCATGCAATGCT 1740
1681 ATCAGGATGTCGGAATTTCTGGCTGCTGATTAAGTTCGTGACCAAGCATGCAATGCT 1740
1741 TGTCTGTGGGTTTGTATGCTGCGGACTGTCGCGCAAGATCAITTTTCAAGTATGTA 1800
1741 TGTCTGTGGGTTTGTATGCTGCGGACTGTCGCGCAAGATCAITTTTCAAGTATGTA 1800
1801 TGATCTCTTCCAAACAGACTCATATATTTATTTCCAAAGGTAATGCTGCTTATTT 1860
1801 TGATCTCTTCCAAACAGACTCATATATTTATTTCCAAAGGTAATGCTGCTTATTT 1860
1861 TCAGCTTTGCGAAGTACGCAAGAGGAGTGTGAGGTCCTATAGTACCAATCCAAATA 1920
1861 TCAGCTTTGCGAAGTACGCAAGAGGAGTGTGAGGTCCTATAGTACCAATCCAAATA 1920
1921 TTTCGACATGCTTCTATTGCGCAACAGTGGGAAACCATCCACTCATATATGACAGTG 1980
1921 TTTCGACATGCTTCTATTGCGCAACAGTGGGAAACCATCCACTCATATATGACAGTG 1980
1981 TGAATGCCACCAACATATATTTTAACTCTCAGCTTTCAAAATACAAACGATCAAGAT 2040
1981 TGAATGCCACCAACATATATTTTAACTCTCAGCTTTCAAAATACAAACGATCAAGAT 2040
```

QY 2041 AAATCAGATACAGTGGAGTGGACACAAGGGAGGACCAAACTGAATTTCTACGGCC 2100  
DB 2041 AAATCAGATACAGTGGAGTGGACACAAGGGAGGACCAAACTGAATTTCTACGGCC 2100  
QY 2101 AGAAGGGTTACGAAATTTAGTTAGTCCCATACACCTTTCCAGAGCGGAATCCCTTT 2150  
DB 2101 AGAAGGGTTACGAAATTTAGTTAGTCCCATACACCTTTCCAGAGCGGAATCCCTTT 2150  
QY 2161 TTGAGGATATCCCAAGAAACCGCTTCCGAAAGTTAAAGACATGATGTTAACTCAA 2220  
DB 2161 TTGAGGATATCCCAAGAAACCGCTTCCGAAAGTTAAAGACATGATGTTAACTCAA 2220  
QY 2221 CAAAGAGAGCCAGAGAGAGTGAATAATCCCTGGTAAATATTTTCACTCCTTCAAAG 2280  
DB 2221 CAAAGAGAGCCAGAGAGAGTGAATAATCCCTGGTAAATATTTTCACTCCTTCAAAG 2280  
QY 2281 AGCCCGAGTTGAGTCTCAATACCTTTGGATTTGCAACTGGAACATGAGACATCACTTTGA 2340  
DB 2281 AGCCCGAGTTGAGTCTCAATACCTTTGGATTTGCAACTGGAACATGAGACATCACTTTGA 2340  
QY 2341 AAGGATACAAATTTGTCAGAGTGGAGTGAATTTCTGATGAACCTCAGAGCATG 2400  
DB 2341 AAGGATACAAATTTGTCAGAGTGGAGTGAATTTCTGATGAACCTCAGAGCATG 2400  
QY 2401 CTAAATATAAATCAAGCTATATAACAGATGAACAAATACACAGTTTGGTGGCTCCAC 2460  
DB 2401 CTAAATATAAATCAAGCTATATAACAGATGAACAAATACACAGTTTGGTGGCTCCAC 2460  
QY 2461 AGGAAACACAGTTTCATATAAGCATCTTCCCAACAGCTTAGGAGTGTCTGAAAGATTGC 2520  
DB 2461 AGGAAACACAGTTTCATATAAGCATCTTCCCAACAGCTTAGGAGTGTCTGAAAGATTGC 2520  
QY 2521 AGAGTTGACTTTCTCGAGTGAAGTGAATAAGTGAATGGTCAACAGGGTCAGAAATC 2580  
DB 2521 AGAGTTGACTTTCTCGAGTGAAGTGAATAAGTGAATGGTCAACAGGGTCAGAAATC 2580  
QY 2581 CACCCCTGACCTTGAGACCAAGCAAGATTTAGAGTGGAACTCACACCCAAAACCA 2640  
DB 2581 CACCCCTGACCTTGAGACCAAGCAAGATTTAGAGTGGAACTCACACCCAAAACCA 2640  
QY 2641 TAGCGGAAATGTGACAAAGAAAGCCCCATCTCTGATTTCCACTGGAAAGCCAGA 2700  
DB 2641 TAGCGGAAATGTGACAAAGAAAGCCCCATCTCTGATTTCCACTGGAAAGCCAGA 2700  
QY 2701 TCACAAAGAAAGAAATACAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2760  
DB 2701 TCACAAAGAAAGAAATACAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2760  
QY 2761 CTGAAATACATAGGCGTTACTGAAGTGTACTTTGGAAGAAAGCTGCAGCATTACACAG 2820  
DB 2761 CTGAAATACATAGGCGTTACTGAAGTGTACTTTGGAAGAAAGCTGCAGCATTACACAG 2820  
QY 2821 ATAGTTACTTGGCTTTTGGCATGGAGAAAGAAAGTATTTCCAGATCTTCTCGAG 2880  
DB 2821 ATAGTTACTTGGCTTTTGGCATGGAGAAAGAAAGTATTTCCAGATCTTCTCGAG 2880  
QY 2881 AAGAGAGTCAATTGAAGACACAAATTTGGCATACTTCACTGATAGCAAAATACTGGAGGC 2940  
DB 2881 AAGAGAGTCAATTGAAGACACAAATTTGGCATACTTCACTGATAGCAAAATACTGGAGGC 2940  
QY 2941 AACTAAAGATACATTTGAGATTCCTCAGATATGTAATAAATTTCTAAATAGCAAGT 3000  
DB 2941 AACTAAAGATACATTTGAGATTCCTCAGATATGTAATAAATTTCTAAATAGCAAGT 3000  
QY 3001 TTGGATTACATCGCGGAAAGTCCCTGCTCACATGCCCTCAGATGATGACCGGATTGTA 3060  
DB 3001 TTGGATTACATCGCGGAAAGTCCCTGCTCACATGCCCTCAGATGATGACCGGATTGTA 3060  
QY 3061 TCAAGAACTGCAAGATATGTTCCCTGGAAGATTTGACAAGAGCTCATTTCAAAAGTGC 3120  
DB 3061 TCAAGAACTGCAAGATATGTTCCCTGGAAGATTTGACAAGAGCTCATTTCAAAAGTGC 3120

QY 3121 GCCATTCTGAGGATATGCGAGTTTGCCTTCTCTTATTTTATATCTCATGAGTGCAGTGC 3180  
DB 3121 GCCATTCTGAGGATATGCGAGTTTGCCTTCTCTTATTTTATATCTCATGAGTGCAGTGC 3180  
QY 3181 AGCCACTGATATATCTCAAGTCTTTGATGAAGTGTGATACAGATCAATCTGGTGTCTTGT 3240  
DB 3181 AGCCACTGATATATCTCAAGTCTTTGATGAAGTGTGATACAGATCAATCTGGTGTCTTGT 3240  
QY 3241 CTGACAGAGAAATCCGAAACACTGGCTACCGAATTCACGAACTGCCGTTAAGTTTCAGG 3300  
DB 3241 CTGACAGAGAAATCCGAAACACTGGCTACCGAATTCACGAACTGCCGTTAAGTTTCAGG 3300  
QY 3301 ATTTGACAGTCTGGAGACATGCTAATAAATTTGCTCAAAAATGCTCTCTGCTGATATCA 3360  
DB 3301 ATTTGACAGTCTGGAGACATGCTAATAAATTTGCTCAAAAATGCTCTCTGCTGATATCA 3360  
QY 3361 CGCAGCTTAAATATTTCCACCAACTCAGGAATCTCTATGATCCCAACCTGCCACCGG 3420  
DB 3361 CGCAGCTTAAATATTTCCACCAACTCAGGAATCTCTATGATCCCAACCTGCCACCGG 3420  
QY 3421 TCACATAAAGTCTAGTAAACAACTGTAAACCACTAATGCTCAAAAATGCTCTCTGCTGATATCA 3480  
DB 3421 TCACATAAAGTCTAGTAAACAACTGTAAACCACTAATGCTCAAAAATGCTCTCTGCTGATATCA 3480  
QY 3481 AGGACAAAAACAATATAGGTTTGAATCATGGGAGAGAAATCGCTTTTAAATATGA 3540  
DB 3481 AGGACAAAAACAATATAGGTTTGAATCATGGGAGAGAAATCGCTTTTAAATATGA 3540  
QY 3541 TTGCTACCAAGTTTCTCATGTTGGCGAGTTGGATGACATAGAAAAACCTTAGGA 3600  
DB 3541 TTGCTACCAAGTTTCTCATGTTGGCGAGTTGGATGACATAGAAAAACCTTAGGA 3600  
QY 3601 AGTTGTTGCTGGAATGACAACTTGAACACATCATAAAGATGCTCAGACAGTGAAG 3660  
DB 3601 AGTTGTTGCTGGAATGACAACTTGAACACATCATAAAGATGCTCAGACAGTGAAG 3660  
QY 3661 CTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAACCTGCCAA 3720  
DB 3661 CTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAACCTGCCAA 3720  
QY 3721 GAGAGTACGAAACCGTTTCCCTCATATGATGCTGAGTGAAGTGAAGTGAAGTGAAG 3780  
DB 3721 GAGAGTACGAAACCGTTTCCCTCATATGATGCTGAGTGAAGTGAAGTGAAGTGAAG 3780  
QY 3781 ACAATTTGAAGTTTGGACCCATTGTACTAGCAACATTTGATTTGTTTACTATATCT 3840  
DB 3781 ACAATTTGAAGTTTGGACCCATTGTACTAGCAACATTTGATTTGTTTACTATATCT 3840  
QY 3841 CATTTTGTGCTGAGCAGTTAATTTGCTTAAGCGGAAGATATTTCCAGAGGAGATAC 3900  
DB 3841 CATTTTGTGCTGAGCAGTTAATTTGCTTAAGCGGAAGATATTTCCAGAGGAGATAC 3900  
QY 3901 ACAAGAGCTAGTCCCAATCGAATCAGAGTATAGAAAGATCTTCAATTTGAAAACCATCTA 3960  
DB 3901 ACAAGAGCTAGTCCCAATCGAATCAGAGTATAGAAAGATCTTCAATTTGAAAACCATCTA 3960  
QY 3961 CCTCAGCATTTCTGAGCATTTTAAACCTCAGTTTCAAGAGATGCTTTGTGATGAT 4020  
DB 3961 CCTCAGCATTTCTGAGCATTTTAAACCTCAGTTTCAAGAGATGCTTTGTGATGAT 4020  
QY 4021 GCTTAGCAGTTTGGCGGAGAGGAAATATCCAGTACCATGCTGTTTGTGCGCATGAA 4080  
DB 4021 GCTTAGCAGTTTGGCGGAGAGGAAATATCCAGTACCATGCTGTTTGTGCGCATGAA 4080  
QY 4081 TATAGCCCACTGACTAGGAATTTTAAACCAACCCACTGAAAACCTTGTGTCGAGCAGC 4140  
DB 4081 TATAGCCCACTGACTAGGAATTTTAAACCAACCCACTGAAAACCTTGTGTCGAGCAGC 4140  
QY 4141 TCTGAACCTGATTTTACTTTTAAAGAAATTTGCTCATGAGCCTGCTCATCTTTTATAAAA 4200  
DB 4141 TCTGAACCTGATTTTACTTTTAAAGAAATTTGCTCATGAGCCTGCTCATCTTTTATAAAA 4200  
QY 4201 GGCTCACTGACAGAGACAGCTGTAAATTTCCCAAGCAATCATTCAGACTAACTTTAT 4260



Db 4201 GGCTCACTGACAGAGAGAGCTGTAAATTTCCACAGCAATCAATGACAGACTACTTTAT 4260  
Qy 4261 TAGGAGAGCCTATGCCAGCTGGGAGTGATTCCTAAGAGGCTCCAGCTCTTTGCAATCCAA 4320  
Db 4261 TAGGAGAGCCTATGCCAGCTGGGAGTGATTCCTAAGAGGCTCCAGCTCTTTGCAATCCAA 4320  
Qy 4321 AGCCCTTTTGTAAAGCTTTTGGACCTTTTCTTTTTCATTTCCCATTTTAACTAGTACTA 4380  
Db 4321 AGCCCTTTTGTAAAGCTTTTGGACCTTTTCTTTTTCATTTCCCATTTTAACTAGTACTA 4380  
Qy 4381 AGTTAACTAGTTATTTCTGCTCTGAGTATTAACGAATTTGGGATCTTAACCTATTTTAA 4440  
Db 4381 AGTTAACTAGTTATTTCTGCTCTGAGTATTAACGAATTTGGGATCTTAACCTATTTTAA 4440  
Qy 4441 TAGATGTTTATTAATATGAGCAATATACCTCTTATTTGACAAATACCTAAATATGAG 4500  
Db 4441 TAGATGTTTATTAATATGAGCAATATACCTCTTATTTGACAAATACCTAAATATGAG 4500  
Qy 4501 TTTTATTAATATTTAAGACTGTAAATGCTTAAACCACTAACTACTGAAGAGCTCAATG 4560  
Db 4501 TTTTATTAATATTTAAGACTGTAAATGCTTAAACCACTAACTACTGAAGAGCTCAATG 4560  
Qy 4561 ATTGACATCTGAATGCTTTGTAATTTATGACTTACGCCCTTGAATGCTATGATTTCA 4620  
Db 4561 ATTGACATCTGAATGCTTTGTAATTTATGACTTACGCCCTTGAATGCTATGATTTCA 4620  
Qy 4621 CGTCAGAGCTCTAATTTCAACAGGCTAGAGTTAGTACTACTTACCAGATGTAATATGTTT 4680  
Db 4621 CGTCAGAGCTCTAATTTCAACAGGCTAGAGTTAGTACTACTTACCAGATGTAATATGTTT 4680  
Qy 4681 TGGAAATGTACATATCAAAAGAGAGTGCCTCATTTTAGAAATGAGTAGTCTGATGGCA 4740  
Db 4681 TGGAAATGTACATATCAAAAGAGAGTGCCTCATTTTAGAAATGAGTAGTCTGATGGCA 4740  
Qy 4741 CTGGCACATTAAGAGTGTCTGTTTAACTACTCAATGGTATATTCAGTAGTCTATCTCT 4800  
Db 4741 CTGGCACATTAAGAGTGTCTGTTTAACTACTCAATGGTATATTCAGTAGTCTATCTCT 4800  
Qy 4801 CTCAGTTGGTTTTGATAGACAGAGGCCACCAACTTTCTTTGTAAGAGCTGGTTAGT 4860  
Db 4801 CTCAGTTGGTTTTGATAGACAGAGGCCACCAACTTTCTTTGTAAGAGCTGGTTAGT 4860  
Qy 4861 AAATATTTGCGAGGCCACCTGTCTTTGTCATACATCTTCTGCTCTCTGTTAGTTGT 4920  
Db 4861 AAATATTTGCGAGGCCACCTGTCTTTGTCATACATCTTCTGCTCTCTGTTAGTTGT 4920  
Qy 4921 TTTTTCCTCAAAACACCTCTAAATATGTAAGAGCTAGTTAGTCTGAGCTGTACAAA 4980  
Db 4921 TTTTTCCTCAAAACACCTCTAAATATGTAAGAGCTAGTTAGTCTGAGCTGTACAAA 4980  
Qy 4981 ACTGCCACAGCAGATGACCTCTAGGCCATCATTTGCCATCATCTGAGAAATATTT 5040  
Db 4981 ACTGCCACAGCAGATGACCTCTAGGCCATCATTTGCCATCATCTGAGAAATATTT 5040  
Qy 5041 TTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 5100  
Db 5041 TTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 5100  
Qy 5101 TGCAGTGGCGCAATCTCAGCTCACTGCAACCTCCGCTCTCCGGGTTCAAGCAGTTCTGTC 5160  
Db 5101 TGCAGTGGCGCAATCTCAGCTCACTGCAACCTCCGCTCTCCGGGTTCAAGCAGTTCTGTC 5160  
Qy 5161 TCAGGCTTTCTGAGTAGCTGGAGCTACAGGTGCATGCGCACACACCTCTGCTAAATTTTGT 5220  
Db 5161 TCAGGCTTTCTGAGTAGCTGGAGCTACAGGTGCATGCGCACACACCTCTGCTAAATTTTGT 5220  
Qy 5221 TTTTATGAGAGAGCGGGGTTCCACCATATTTGGTCAAGCTTATCTGAACTCTGACCTC 5280  
Db 5221 TTTTATGAGAGAGCGGGGTTCCACCATATTTGGTCAAGCTTATCTGAACTCTGACCTC 5280  
Qy 5281 AGGTGATCCACTGCTCTGCTCTCCCAAGTGTGAGTTACAGGCATAGCCAGTGCAC 5340

Db 5281 AGGTGATCCACTGCTCTGCTCTCCCAAGTGTGAGTTACAGGCATAGCCAGTGCAC 5340  
Qy 5341 CGAGCCGAGAAATAGTATTTTATGTTGTTTAAACCTTGGCTTACCCATATTTTATG 5400  
Db 5341 CGAGCCGAGAAATAGTATTTTATGTTGTTTAAACCTTGGCTTACCCATATTTTATG 5400  
Qy 5401 TCATATAACAATCGAATTTCTGAAGAGCAGATTTCCATGAGTAACTCTGACAGATATTTAG 5460  
Db 5401 TCATATAACAATCGAATTTCTGAAGAGCAGATTTCCATGAGTAACTCTGACAGATATTTAG 5460  
Qy 5461 ATCATGATCTCAACAATATCTCCCAATGCAATACATCTTTTGTACAAAGAACTTCAA 5520  
Db 5461 ATCATGATCTCAACAATATCTCCCAATGCAATACATCTTTTGTACAAAGAACTTCAA 5520  
Qy 5521 ATGTAAATACTGTTGTTGCTGTAAGAGTCTGTTATTTCAAAAACCTCAATCAA 5580  
Db 5521 ATGTAAATACTGTTGTTGCTGTAAGAGTCTGTTATTTCAAAAACCTCAATCAA 5580  
Qy 5581 AAGTTAAATTTTGAATA 5597  
Db 5581 AAGTTAAATTTTGAATA 5597

## RESULT 5

US-09-636-596C-4  
; Sequence 4, Application US/09636596C  
; Patent No. 6770468  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM  
; TITLE OF INVENTION: PHOSPHODIESTER ALPHA-GLUCANase OF THE LYSOSOMAL TARGETING PATHWAY  
; FILE REFERENCE: 10929-0001-77  
; CURRENT APPLICATION NUMBER: US/09/636,596C  
; CURRENT FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 5597  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-636-596C-4

Query Match 100.0%; Score 5597; DB 4; Length 5597;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 CGAGCCGAGCGGGGCTCCGTCGCGGAGCTGCAATGAGCGCGCCCGAGGCTGTGACC 60  
Db 1 CGAGCCGAGCGGGGCTCCGTCGCGGAGCTGCAATGAGCGCGCCCGAGGCTGTGACC 60  
Qy 61 TCGGCGCGCGCGCGAGCGGGGCGCCCTGTAATGGCGGCTCCGTCGAGCGCGCGCGCGC 120  
Db 61 TCGGCGCGCGCGCGCGAGCGGGGCGCCCTGTAATGGCGGCTCCGTCGAGCGCGCGCGC 120  
Qy 121 GCGCGCTCAGGCTCCCTCGCGGGGCTGCGCGGTGGAAGGGGTGATGCTGTTCAAGTCC 180  
Db 121 GCGCGCTCAGGCTCCCTCGCGGGGCTGCGCGGTGGAAGGGGTGATGCTGTTCAAGTCC 180  
Qy 181 TGCAGAGACAAACCTATACCTGCTGCTGTCGAGATGAGGCTCTACGCTGCTCTTGG 240  
Db 181 TGCAGAGACAAACCTATACCTGCTGCTGTCGAGATGAGGCTCTACGCTGCTCTTGG 240  
Qy 241 GCGTCTGTTGTCACCATCTGCTCCGCGCTTCCAGTTCGAGAGAGTGGTTCGGAATGGAGCC 300  
Db 241 GCGTCTGTTGTCACCATCTGCTCCGCGCTTCCAGTTCGAGAGAGTGGTTCGGAATGGAGCC 300  
Qy 301 GAGATCAATACCATGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 360  
Db 301 GAGATCAATACCATGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 360  
Qy 361 AGAATCGGCTTGTCTGCCCATGCGGATGAGTGTGTTTACACCTGGGTGATGCGACAG 420





QY	2581	CACCCCTGGAGCTTGGAGACCAACAGATTTAGAGTGGAACTCACACCCAAAAACCA	2640	Db	3661	CTGTTCTCAGGACTTCTATGAATCCATGTTTCCCATACCTTCCCAATTTGAACATGCCAA	3720
Db	2581	CACCCCTGGAGCTTGGAGACCAACAGATTTAGAGTGGAACTCACACCCAAAAACCA	2640	QY	3721	GAGAGTATCGAAAACCGTTTCTTATGTCATGATGAGCTGCAGGAATGGAGGGCTTATCCAG	3780
QY	2641	TAGGCGGAAATGTGACAAAGAAAGACCCCATCTCTGATTGTTCCACTGCGAAGCCAGA	2700	Db	3721	GAGAGTATCGAAAACCGTTTCTTATGTCATGATGAGCTGCAGGAATGGAGGGCTTATCCAG	3780
Db	2641	TAGGCGGAAATGTGACAAAGAAAGACCCCATCTCTGATTGTTCCACTGCGAAGCCAGA	2700	QY	3781	ACAAATGGAAGTTTGGACCCATTTGTCATGATAGCAACATTTGATTTATGTTATATTTCT	3840
QY	2701	TGACAAAAAGAAAGAAATTCACAGGAAAGAAAAAGAGAACAGTAGAATGAGGAAATG	2760	Db	3781	ACAAATGGAAGTTTGGACCCATTTGTCATGATAGCAACATTTGATTTATGTTATATTTCT	3840
Db	2701	TGACAAAAAGAAAGAAATTCACAGGAAAGAAAAAGAGAACAGTAGAATGAGGAAATG	2760	QY	3841	CATTTTTCCTCAGCAGATTAATTGCACTTAAGCGGAAGATATTTCCCAAGAGGAGGATAC	3900
QY	2761	CTGAAATACATAGGCGCTTACTGAGTGTACTTGGNAGAAAGCTGCAGCATTTACACAG	2820	Db	3841	CATTTTTCCTCAGCAGATTAATTGCACTTAAGCGGAAGATATTTCCCAAGAGGAGGATAC	3900
Db	2761	CTGAAATACATAGGCGCTTACTGAGTGTACTTGGNAGAAAGCTGCAGCATTTACACAG	2820	QY	3901	ACAAAGAGCTAGTCCCAATCGAATCAGAGTATAGAGATCTTCATTTTGAAACCATCTA	3960
QY	2821	ATAGTTACTTGGGCTTTTGGCATGGGAGAAAAAAGTATTTCCAAAGATCTTCTCGACG	2880	Db	3901	ACAAAGAGCTAGTCCCAATCGAATCAGAGTATAGAGATCTTCATTTTGAAACCATCTA	3960
Db	2821	ATAGTTACTTGGGCTTTTGGCATGGGAGAAAAAAGTATTTCCAAAGATCTTCTCGACG	2880	QY	3961	CCTCAGCATTTACTCAGCAGATTTTAAACTCTCAGATTCACAGAGATGTCTTTGTGATGTAT	4020
QY	2881	AAGAGAGTCAATGAAACACAAATGGCATACTTCACTGATAGCAAAAAATCTGGGAGGC	2940	Db	3961	CCTCAGCATTTACTCAGCAGATTTTAAACTCTCAGATTCACAGAGATGTCTTTGTGATGTAT	4020
Db	2881	AAGAGAGTCAATGAAACACAAATGGCATACTTCACTGATAGCAAAAAATCTGGGAGGC	2940	QY	4021	GCTTAGCAGTTTGGGCCGGAAGGAAAAATATCCAGTACCATGCTGTTTGTGGCATGAA	4080
QY	2941	AACTAAGAGATACATTTCCAGATTCCTTCAGATATGTAAATAAAATTTCTAAATAGCAAGT	3000	Db	4021	GCTTAGCAGTTTGGGCCGGAAGGAAAAATATCCAGTACCATGCTGTTTGTGGCATGAA	4080
Db	2941	AACTAAGAGATACATTTCCAGATTCCTTCAGATATGTAAATAAAATTTCTAAATAGCAAGT	3000	QY	4081	TATAGCCCATGACCTAGGAATTTTAAACCAACCCACTGAAAACTTGTGTGTCGAGCAGC	4140
QY	3001	TTGGATTCACATCGCGGAAATGCTGCTCATGCTCATGCTCATGATGACCGGATTTGTTA	3060	Db	4081	TATAGCCCATGACCTAGGAATTTTAAACCAACCCACTGAAAACTTGTGTGTCGAGCAGC	4140
Db	3001	TTGGATTCACATCGCGGAAATGCTGCTCATGCTCATGCTCATGATGACCGGATTTGTTA	3060	QY	4141	TCGAACTGATTTTACTTTTAAAGAAATTTGCTCATGACCTGTCTATCTCTTTTATAAAA	4200
QY	3061	TGCAAGAACTGCAAGATGTTTCCCTGAAGAAATTTGACAGAGCTCATTTTACAAAGTGC	3120	Db	4141	TCGAACTGATTTTACTTTTAAAGAAATTTGCTCATGACCTGTCTATCTCTTTTATAAAA	4200
Db	3061	TGCAAGAACTGCAAGATGTTTCCCTGAAGAAATTTGACAGAGCTCATTTTACAAAGTGC	3120	QY	4201	GGCTCAGTACAGAGACAGCTGTTAAATTTCCACAGCAATCATTTGAGAGCTAATCTTTAT	4260
QY	3121	GCCATTCGAGGATATGAGTTCCTCTCTCTTATTTTATTTATTCATGAGTGCAGTGC	3180	Db	4201	GGCTCAGTACAGAGACAGCTGTTAAATTTCCACAGCAATCATTTGAGAGCTAATCTTTAT	4260
Db	3121	GCCATTCGAGGATATGAGTTCCTCTCTCTTATTTTATTTATTCATGAGTGCAGTGC	3180	QY	4261	TAGGAGAGAGCTTATGCCAGCTGGGAGTCAATGCTAAGAGGCTCCAGTCTTTTGCAATCCAA	4320
QY	3181	AGCCACTGAATATATCTCAAGTCTTTGATGAGTGTATACAGATCAATCTGGTGTCTGT	3240	Db	4261	TAGGAGAGAGCTTATGCCAGCTGGGAGTCAATGCTAAGAGGCTCCAGTCTTTTGCAATCCAA	4320
Db	3181	AGCCACTGAATATATCTCAAGTCTTTGATGAGTGTATACAGATCAATCTGGTGTCTGT	3240	QY	4321	AGCCCTTTTGCTAAAGTTTGCACCTTTTTCATTTCCCATTTTAAAGTAGTACTA	4380
QY	3241	CTGACAGAGAAATCCGAAACATCGGCTACCAAGATTCACGAACTGCGCTTAAGTTTGAGG	3300	Db	4321	AGCCCTTTTGCTAAAGTTTGCACCTTTTTCATTTCCCATTTTAAAGTAGTACTA	4380
Db	3241	CTGACAGAGAAATCCGAAACATCGGCTACCAAGATTCACGAACTGCGCTTAAGTTTGAGG	3300	QY	4381	AGTTAACTAGTTTATCTTCTGCTTCTGAGTATACGAAATGGGATGTCTAAACCTATTTTA	4440
QY	3301	ATTTGACAGGCTGGAACACATGCTATATAAATTTGCTCAAAAATGCTTCTGCTGATATCA	3360	Db	4381	AGTTAACTAGTTTATCTTCTGCTTCTGAGTATACGAAATGGGATGTCTAAACCTATTTTA	4440
Db	3301	ATTTGACAGGCTGGAACACATGCTATATAAATTTGCTCAAAAATGCTTCTGCTGATATCA	3360	QY	4441	TAGATGTTTATTAATAATGCAGCAATATCACCTCTTATTGACAAATACCTAAATATGAG	4500
QY	3361	CGCAGTAAATAATTTCCACCACTCAGGAATCCTACTATGATCCCAACCTGCCACCGG	3420	Db	4441	TAGATGTTTATTAATAATGCAGCAATATCACCTCTTATTGACAAATACCTAAATATGAG	4500
Db	3361	CGCAGTAAATAATTTCCACCACTCAGGAATCCTACTATGATCCCAACCTGCCACCGG	3420	QY	4501	TTTTTATTAATTTTAAAGCTGTAATGCTTAAACCACTAATCTACTGAGAGCTCAATG	4560
QY	3421	TCACATAAAGCTAGTAAACAACTGTAAACAGATTAACAGTAATCCAAAGACATATA	3480	Db	4501	TTTTTATTAATTTTAAAGCTGTAATGCTTAAACCACTAATCTACTGAGAGCTCAATG	4560
Db	3421	TCACATAAAGCTAGTAAACAACTGTAAACAGATTAACAGTAATCCAAAGACATATA	3480	QY	4561	ATTGACATCTGAAATGCTTTGTAATTTATGACTTCAGCCCTCAGCAATGCTATGATTTCA	4620
QY	3481	AGGACAAAACAAATATAGGTTTGAATTCATGGGAGAGAGAAATCGCTTTTAAATGA	3540	Db	4561	ATTGACATCTGAAATGCTTTGTAATTTATGACTTCAGCCCTCAGCAATGCTATGATTTCA	4620
Db	3481	AGGACAAAACAAATATAGGTTTGAATTCATGGGAGAGAGAAATCGCTTTTAAATGA	3540	QY	4621	CGTGAGGCTAAATTTCAACAGGCTAGAGTACTACTTACCAGATGTAATATGTTTT	4680
QY	3541	TTCTGTACCAAGCTTCTCATGTGTTGGCCAGTTGGATGACATAAGAAAAAACCCCTAGGA	3600	Db	4621	CGTGAGGCTAAATTTCAACAGGCTAGAGTACTACTTACCAGATGTAATATGTTTT	4680
Db	3541	TTCTGTACCAAGCTTCTCATGTGTTGGCCAGTTGGATGACATAAGAAAAAACCCCTAGGA	3600	QY	4681	TGGAATGTACATATTCACACAGAGTGCCTCATTTTGAAGAAATGAGTAGTCTGATGCA	4740
QY	3601	AGTTTGTGCTGCAATGACAACTTACCAACATCATTAAGATGCTCAGACAGTGAAGG	3660	Db	4681	TGGAATGTACATATTCACACAGAGTGCCTCATTTTGAAGAAATGAGTAGTCTGATGCA	4740
Db	3601	AGTTTGTGCTGCAATGACAACTTACCAACATCATTAAGATGCTCAGACAGTGAAGG	3660	QY	4741	CTGGCACATTCAGTGGTGTCTTGTTTAATACTCATTTGTTATATTCAGTAGTACTTCTCT	4800
QY	3661	CTGTTCTCAGGAGCTTCTATGAATCCATGTTTCCCATACCTTCCCAATTTGAACTGCCAA	3720				



QY	1125	AGTCGTTTTGAAGATAACGAAAGAACTGAGGTACTCATTCGGATCTATCGAGAGCCATGCA	1184
DB	967	AGTCGTTTTGAAGATAACGAAAGAACTGAGGTACTCATTCGGATCTATCGAGAGCCATGCA	1026
QY	1185	CCATGGGTTCCGAATATTTTCATTTGTACCAACGGGAGATTCCTATCCTGGCTGAACCTT	1244
DB	1027	CCATGGGTTCCGAATATTTTCATTTGTACCAACGGGAGATTCCTATCCTGGCTGAACCTT	1086
QY	1245	GCAATCTCCAGTGAACAAATAGTAAACACACAGAGATGTTTTTCGAATTTGAGCCATTG	1304
DB	1087	GCAATCTCCAGTGAACAAATAGTAAACACACAGAGATGTTTTTCGAATTTGAGCCATTG	1146
QY	1305	CCTACCTTTAGTTCACCTGCTATTGAAAGTCACATTCATCGCATCGAAGGCTGTCAG	1364
DB	1147	CCTACCTTTAGTTCACCTGCTATTGAAAGTCACATTCATCGCATCGAAGGCTGTCAG	1206
QY	1365	AAGTTTATTTACCTAAATGATGTCTATGTTTGGGAAGATGTCCTGGCAGATGATTTT	1424
DB	1207	AAGTTTATTTACCTAAATGATGTCTATGTTTGGGAAGATGTCCTGGCAGATGATTTT	1266
QY	1425	TACAGTCTCCAAAGGCCAGAGGTTTATTTGATCGCTGTCGCAAACTGTCGCGAG	1484
DB	1267	TACAGTCTCCAAAGGCCAGAGGTTTATTTGATCGCTGTCGCAAACTGTCGCGAG	1326
QY	1485	GGCTGCCAGGTTCTGTGATTAAGGATGGCTATTGTGACAAAGGCTTGTATAATTGAGCC	1544
DB	1327	GGCTGCCAGGTTCTGTGATTAAGGATGGCTATTGTGACAAAGGCTTGTATAATTGAGCC	1386
QY	1545	TCCGATTTGGGATGTTGGGATTTGCTCTGGAAACAGTGGAGGAGTCTGCTATTTGAGGA	1604
DB	1387	TCCGATTTGGGATGTTGGGATTTGCTCTGGAAACAGTGGAGGAGTCTGCTATTTGAGGA	1446
QY	1605	GGTGAGGATCTGGGAGTATTTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAAC	1664
DB	1447	GGTGGAGGATCTGGGAGTATTTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAAC	1506
QY	1665	AGTGTCCTTACTGTAATACGGGATGTCGGAATTTCTGGCTGCTGTATTAAGTTCTGTGAC	1724
DB	1507	AGTGTCCTTACTGTAATACGGGATGTCGGAATTTCTGGCTGCTGTATTAAGTTCTGTGAC	1566
QY	1725	CAAGCATGCAATGCTTCTGCTGGGTTTTCATCTGCGCATCTGCGGCAAGATCATTTT	1784
DB	1567	CAAGCATGCAATGCTTCTGCTGGGTTTTCATCTGCGCATCTGCGGCAAGATCATTTT	1626
QY	1785	CATGAATTTGATAAAGTGATCTTCTCCAAACAGACTCACTATATTTATTTCCAAAAGGT	1844
DB	1627	CATGAATTTGATAAAGTGATCTTCTCCAAACAGACTCACTATATTTATTTCCAAAAGGT	1686
QY	1845	GAATGCTGCTTATTTTTCAGTTTTCAGAGTACCAAGAGAGGATTTGAAGTGCCTAT	1904
DB	1687	GAATGCTGCTTATTTTTCAGTTTTCAGAGTACCAAGAGAGGATTTGAAGTGCCTAT	1746
QY	1905	AGTGACATCAATAAATTCGACATGCTTCTATTGCCAACAGTGGAAGAACCATCCACCTC	1964
DB	1747	AGTGACATCAATAAATTCGACATGCTTCTATTGCCAACAGTGGAAGAACCATCCACCTC	1806
QY	1965	ATAATGACAGTGAATGCAACCAATAGATTTTATTTTTCAGTTTTCAAAATACA	2024
DB	1807	ATAATGACAGTGAATGCAACCAATAGATTTTATTTTTCAGTTTTCAAAATACA	1866
QY	2025	AACGATGAAGAGTTTCAAAATGCAATACAGTGGAGTGACACAAAGGGAGGACCAAAA	2084
DB	1867	AACGATGAAGAGTTTCAAAATGCAATACAGTGGAGTGACACAAAGGGAGGACCAAAA	1926
QY	2085	CTGAATTTACGGCCCAAGAGGTTTACGAAATTTAGTTAGTCCCATTAACACTTTTCCA	2144
DB	1927	CTGAATTTACGGCCCAAGAGGTTTACGAAATTTAGTTAGTCCCATTAACACTTTTCCA	1986
QY	2145	GAGCGGAAATCTTTTGGAGATTTTCCCAAGAAACGCTTCCCGAAGTTTAAAGA	2204
DB	1987	GAGCGGAAATCTTTTGGAGATTTTCCCAAGAAACGCTTCCCGAAGTTTAAAGA	2046

QY	2205	CATGATGTTTAACTCAACAGGAGAGCCCGAGGAGAGGTGAAATTTCCCTCGTAAATATT	2264
DB	2047	CATGATGTTTAACTCAACAGGAGAGCCCGAGGAGAGGTGAAATTTCCCTCGTAAATATT	2106
QY	2265	TCACCTCTTTCCAAAAGAGCCCGAGTTGAGTCTCAATACCTTTGGATTTTCAACTGGAACAT	2324
DB	2107	TCACCTCTTTCCAAAAGAGCCCGAGTTGAGTCTCAATACCTTTGGATTTTCAACTGGAACAT	2166
QY	2325	GGAGACATCACTTTGAAAGGATACAAATTTGTCCAAGTCAGCTTGCAGATCAATTTCTG	2384
DB	2167	GGAGACATCACTTTGAAAGGATACAAATTTGTCCAAGTCAGCTTGCAGATCAATTTCTG	2226
QY	2385	ATGAATCTCACAGCATGTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	2444
DB	2227	ATGAATCTCACAGCATGTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	2286
QY	2445	AGTTTGGTGGTCCACAGGAAAAACAGGTTTCAATAAAGATCTTTGCCAAAAACAGTTAGGA	2504
DB	2287	AGTTTGGTGGTCCACAGGAAAAACAGGTTTCAATAAAGATCTTTGCCAAAAACAGTTAGGA	2346
QY	2505	GTGCTCTGAAAGATTCAGAGGTTGACTTTTCTGTCAGTGAGTGTAAAGTGAATGCTCAT	2564
DB	2347	GTGCTCTGAAAGATTCAGAGGTTGACTTTTCTGTCAGTGAGTGTAAAGTGAATGCTCAT	2406
QY	2565	GACCAAGGTCAGAATCCACCCCTGGAATTTGGAGACCAACAGCAAGATTTAGAGTGAAACT	2624
DB	2407	GACCAAGGTCAGAATCCACCCCTGGAATTTGGAGACCAACAGCAAGATTTAGAGTGAAACT	2466
QY	2625	CACACCCAAAAACCATAGGCGGAAATGTGACAAAGAAAGAGCCCACTCTGATGCTT	2684
DB	2467	CACACCCAAAAACCATAGGCGGAAATGTGACAAAGAAAGAGCCCACTCTGATGCTT	2526
QY	2685	CCACTGGAAAGCCAGATGACAAAAGAAAGAAATTCACAGGAAAGAAAGAAAGAAAGT	2744
DB	2527	CCACTGGAAAGCCAGATGACAAAAGAAAGAAATTCACAGGAAAGAAAGAAAGAAAGT	2586
QY	2745	AGAATGGAGAAATGCTGAAAAATCAGATAGGCGTTACTGAAGTGTACTTTGGAGAAAG	2804
DB	2587	AGAATGGAGAAATGCTGAAAAATCAGATAGGCGTTACTGAAGTGTACTTTGGAGAAAG	2646
QY	2805	CTGCAGCATTTACACAGATAGTTACTTTGGGCTTTTTCCTGCGGAGAAAGAAAGTATTTTC	2864
DB	2647	CTGCAGCATTTACACAGATAGTTACTTTGGGCTTTTTCCTGCGGAGAAAGAAAGTATTTTC	2706
QY	2865	CAAGATCTTCTCGAGAAAGAGAGTCAATTTGAAGACAAATTTGGCATCTTCACTGATAGC	2924
DB	2707	CTAGATCTTCTCGAGAAAGAGAGTCAATTTGAAGACAAATTTGGCATCTTCACTGATAGC	2766
QY	2925	AAAAATCTGGGAGGCACTAAAGATACATTTGAGATTTCCCTCAGATATGTAATAAATAA	2984
DB	2767	AAAAATCTGGGAGGCACTAAAGATACATTTGAGATTTCCCTCAGATATGTAATAAATAA	2826
QY	2985	ATTCTAAATAGCAAGTTTGGATTCATCTCGGGAAGTCCCTGCTCACTGCTCAATG	3044
DB	2827	ATTCTAAATAGCAAGTTTGGATTCATCTCGGGAAGTCCCTGCTCACTGCTCACTG	2886
QY	3045	ATTGACCGGATTTGATTCGAAGATCAAGATATGTTCCCTGAGATTTTGAAGATTTGACAGAG	3104
DB	2887	ATTGACCGGATTTGATTCGAAGATCAAGATATGTTCCCTGAGATTTTGAAGATTTGACAGAG	2946
QY	3105	TCATTTCAAAAGTGGCCATTTGAGGATATGCAAGTTTGGCTTCTCTTTTATTTAT	3164
DB	2947	TCATTTCAAAAGTGGCCATTTGAGGATATGCAAGTTTGGCTTCTCTTTTATTTAT	3006
QY	3165	CTCATGATGTCAGTCCAGCCACTGAAATATATCTCAAGTCTTTGATGAAGTTGATACAGAT	3224
DB	3007	CTCATGATGTCAGTCCAGCCACTGAAATATATCTCAAGTCTTTGATGAAGTTGATACAGAT	3066
QY	3225	CAATCTGCTGCTTGTCTGACAGAGAAATCCGAACACTGGCTACAGAAATTCAGGAACTG	3284
DB	3067	CAATCTGCTGCTTGTCTGACAGAGAAATCCGAACACTGGCTACAGAAATTCAGGAACTG	3126
QY	3285	CGTTAAGTTTGCAGGATTTGACAGGTTCTGGAAACACTGCTAAATAATTTGCTCAAAAATG	3344

```
Db 3127 CCGTTAAGTTTTCAGAGTTTGACAGTCTGGAACACATGCTAAATAATGCTCAAAAATG 3186
Qy 3345 CTTCTGCTGTATATCAGCAGCTAAATAATATTCACCAACTCAGGAATCCTACTATGAT 3404
Db 3187 CTTCTGCTGTATATCAGCAGCTAAATAATATTCACCAACTCAGGAATCCTACTATGAT 3246
Qy 3405 CCCAACCTGCCACCCTGCTCACTAAAAGTGTAGTAACAAATGTAAACCAAGTAACTGACAAA 3464
Db 3247 CCCAACCTGCCACCCTGCTCACTAAAAGTGTAGTAACAAATGTAAACCAAGTAACTGACAAA 3306
Qy 3465 ATCCCAAGAGCATATAAGGACAAAACAAATATAGTTTGAATCATGCGGAGAGAGAA 3524
Db 3307 ATCCCAAGAGCATATAAGGACAAAACAAATATAGTTTGAATCATGCGGAGAGAGAA 3366
Qy 3525 ATCGCTTTTAAATGATTGTAACCAACGTTTCTCATGTGTTGGCGAGTTGGATGACATA 3584
Db 3367 ATCGCTTTTAAATGATTGTAACCAACGTTTCTCATGTGTTGGCGAGTTGGATGACATA 3426
Qy 3585 AGAAAAACCTTAGAAGTTTGTTCCTGTAATGACACATGACCACATATAAAGAT 3644
Db 3427 AGAAAAACCTTAGAAGTTTGTTCCTGTAATGACACATGACCACATATAAAGAT 3486
Qy 3645 GCTCAGACAGTGAAGGCTGTTCTCAGGAGTCTTATGAATCCATGTTCCCATACCTTCC 3704
Db 3487 GCTCAGACAGTGAAGGCTGTTCTCAGGAGTCTTATGAATCCATGTTCCCATACCTTCC 3546
Qy 3705 CAATTGTAACCTCCAGAGAGTATCGAAACCGTTTCCCTTCATATGATGATGATGATGATGAT 3764
Db 3547 CAATTGTAACCTCCAGAGAGTATCGAAACCGTTTCCCTTCATATGATGATGATGATGATGAT 3606
Qy 3765 TCGAGGGCTTATCGAGACAAATGAAGTTTGGAGCCCATGTTGTAAGTGAAGTGAAGTGAAGT 3824
Db 3607 TCGAGGGCTTATCGAGACAAATGAAGTTTGGAGCCCATGTTGTAAGTGAAGTGAAGTGAAGT 3666
Qy 3825 ATGTTTACTATATCTCAATTTTCTGAGCAGTTTAAATGCACTTAAGCGGAAGATATTT 3884
Db 3667 ATGTTTACTATATCTCAATTTTCTGAGCAGTTTAAATGCACTTAAGCGGAAGATATTT 3726
Qy 3885 CCCAGAGGAGTATACAAAGAGCTAGTCCCAATCGAATCAGATATAGAGATC 3941
Db 3727 CCCAGAGGAGTATACAAAGAGCTAGTCCCAATCGAATCAGATATAGAGATC 3783

RESULT 7
US-09-636-077A-20
; Sequence 20, Application US/09636077A
; Patent No. 6537785
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS OF TREATING LYXOSOMAL STORAGE DISEASE
; FILE REFERENCE: 195612US0
; CURRENT APPLICATION NUMBER: US/09/636,077A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 3783
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-636-077A-20

Query Match 67.2%; Score 3761; DB 4; Length 3783;
Best Local Similarity 99.7%; Pred No 0;
Matches 3767; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 165 ATGCTGTTCAAGCTCTCTGAGAGACAAACCTATACCTGCTGTGCCACAGGTATGGGCTC 224
Db 7 ATGGGGTTCAAGCTCTCTGAGAGACAAACCTATACCTGCTGTGCCACAGGTATGGGCTC 66
Qy 225 TACGTGTGCTTCTTGGGGTCTGTTGCACCATGCTCTCGGCTTCCAGTTCGGAGAGTG 284
```

```
Db 67 TAGCTGTGCTTCTTGGCGCTGTTGTACCATGCTCTCGGCTTCCAGTTTCGGAGAGTG 126
Qy 285 GTTCTGGAATGGAGCCGAGATCAATACCATGTTTGTGTTGATTCCTATAGACAAATATT 344
Db 127 GTTCTGGAATGGAGCCGAGATCAATACCAATGTTTGTGTTGATTCCTATAGACAAATATT 186
Qy 345 GCTGGAAAGTCTTTTCAGAAATCGGCTTGTGTCCTCATGCGGATTCAGCTGTTTACACC 404
Db 187 GCTGGAAAGTCTTTTCAGAAATCGGCTTGTGTCCTCATGCGGATTCAGCTGTTTACACC 246
Qy 405 TGGGTGAATGGCACAGATCTTGAACCTACTGAAGAACTACAGCAGCTCAGAGAACAGATG 464
Db 247 TGGGTGAATGGCACAGATCTTGAACCTACTGAAGAACTACAGCAGCTCAGAGAACAGATG 306
Qy 465 GAGGAGAGCAGAAAACCAATGAGAGAAATCTCTTGGGAAAAACACAAACGGAACCTACTAAG 524
Db 307 GAGGAGAGCAGAAAACCAATGAGAGAAATCTCTTGGGAAAAACACAAACGGAACCTACTAAG 366
Qy 525 AAGAGTGAGAGCAGTATGAGTGTGTTGCTTAACACATCTGCAATTAAGGTGCCAATGTTGA 584
Db 367 AAGAGTGAGAGCAGTATGAGTGTGTTGCTTAACACATCTGCAATTAAGGTGCCAATGTTGA 426
Qy 585 CTGAGCCCAAGCTTGCAGCAACATCACCTGAAGAGAGTGCATCTCTTTATCCTTCT 644
Db 427 CTGAGCCCAAGCTTGCAGCAACATCACCTGAAGAGAGTGCATCTCTTTATCCTTCT 486
Qy 645 TTTCAATCTGCCAGTGACATTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGTC 704
Db 487 TTTCAATCTGCCAGTGACATTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGTC 546
Qy 705 TCAGTTGTTGTTTGTGACAGTACTAAGGATGTTGAAGATGTCGCCACCTGAGCTGCTTAAA 764
Db 547 TCAGTTGTTGTTTGTGACAGTACTAAGGATGTTGAAGATGTCGCCACCTGAGCTGCTTAAA 606
Qy 765 GGAATAAGCAGACAGATGATGAGGGGTACTTGACAAACAGATAAAGAGTCCCTGGA 824
Db 607 GGAATAAGCAGACAGATGATGAGGGGTACTTGACAAACAGATAAAGAGTCCCTGGA 666
Qy 825 TTAGTCTAATGCAAGTATGCGCTTTCCTGAGTGGATTTCCACCACATTTCAAGGAAACA 884
Db 667 TTAGTCTAATGCAAGTATGCGCTTTCCTGAGTGGATTTCCACCACATTTCAAGGAAACA 726
Qy 885 AATCAACTAAAAACAAATTTGCCAGAAAAATCTTCTCTAAAGTCAAACTGTTTGCAGTTG 944
Db 727 AATCAACTAAAAACAAATTTGCCAGAAAAATCTTCTCTAAAGTCAAACTGTTTGCAGTTG 786
Qy 945 TATTCAGAGCCAGTGTAGCGCTTCTAAACTGAATAACCCCAAGATTTTCAAGATTTG 1004
Db 787 TATTCAGAGCCAGTGTAGCGCTTCTAAACTGAATAACCCCAAGATTTTCAAGATTTG 846
Qy 1005 AATAAGCAAACTAAGAGAACATGACATTTGATGGAAGAAAGTACCACATTAAGTCTGCA 1064
Db 847 AATAAGCAAACTAAGAGAACATGACATTTGATGGAAGAAAGTACCACATTAAGTCTGCA 906
Qy 1065 TATTTATATGGGATCTGAGCGCCATCAGCAGCTCTAAGCAGGATGAAGACATCTCTGCC 1124
Db 907 TATTTATATGGGATCTGAGCGCCATCAGCAGCTCTAAGCAGGATGAAGACATCTCTGCC 966
Qy 1125 AGTCGTTTTCAGATAAACGAAGAACTGAGTCTCATTTGCGATCTATCGAGAGCGCATGA 1184
Db 967 AGTCGTTTTCAGATAAACGAAGAACTGAGTCTCATTTGCGATCTATCGAGAGCGCATGA 1026
Qy 1185 CCATGGGTTCCGAAATATTTTCAATGTACCAACGGGCAGATTCATCTCGGCTGAAACCTT 1244
Db 1027 CCATGGGTTCCGAAATATTTTCAATGTACCAACGGGCAGATTCATCTCGGCTGAAACCTT 1086
Qy 1245 GACAACTCTGAGTGACATAGTAACACACCAGGATGTTTTTGAATTTGAGCCACTTG 1304
Db 1087 GACAACTCTGAGTGACATAGTAACACACCAGGATGTTTTTGAATTTGAGCCACTTG 1146
Qy 1305 CCTACCTTTAGTTCACTGCTATTTGAAGTCACTTCATCGCATCGAAGGGTGTGCCAG 1364
```



Db 1147 CCTACCTTTAGTTTACCTGCTATTGAAAGTCACGTTTCATCGCATCGAAGGCGCTGCCAG 1206  
Qy 1365 AGTTTATTTACCTAAATGATGATGCTCATGTTTGGGAGGATGTCGGCCAGATGATTTT 1424  
Db 1207 AAGTTTATTTACCTAAATGATGATGCTCATGTTTGGGAGGATGTCGGCCAGATGATTTT 1266  
Qy 1425 TACAGTCACTCCAAAGGCCAGAGGTTTATTTGACATGGCCCTGTGCCAAACTGTGCCGAG 1484  
Db 1267 TACAGTCACTCCAAAGGCCAGAGGTTTATTTGACATGGCCCTGTGCCAAACTGTGCCGAG 1326  
Qy 1485 GGCTGCCAGGTTCTGCTGAATTAAGGATGGCTATTGTGACAGGCTTGTATTAATTCAGCC 1544  
Db 1327 GGCTGCCAGGTTCTGCTGAATTAAGGATGGCTATTGTGACAGGCTTGTATTAATTCAGCC 1386  
Qy 1545 TCGCATTTGGGATGTTGGGATTTGCTCTGAAACAGTGGAGGAGTCTGCTATTTGCAAGG 1604  
Db 1387 TCGCATTTGGGATGTTGGGATTTGCTCTGAAACAGTGGAGGAGTCTGCTATTTGCAAGG 1446  
Qy 1605 GGTGAGGTACTGGAGTATTGGAGTTGGACACCCCTGGGAGTTTGGTGGAGGATTAAC 1664  
Db 1447 GGTGAGGTACTGGAGTATTGGAGTTGGACACCCCTGGGAGTTTGGTGGAGGATTAAC 1506  
Qy 1665 AGTGCTCTTACTGTAATCAGGATGTCGGAATTCCTGGCTCGCTGATTAAGTTCTGTGAC 1724  
Db 1507 AGTGCTCTTACTGTAATCAGGATGTCGGAATTCCTGGCTCGCTGATTAAGTTCTGTGAC 1566  
Qy 1725 CAAGCATGCAATGCTTCTGCTGGGTTGATGCTGGCGACTGTGGCGAAGATCATTTT 1784  
Db 1567 CAAGCATGCAATGCTTCTGCTGGGTTGATGCTGGCGACTGTGGCGAAGATCATTTT 1626  
Qy 1785 CATGAATTTGATTAAGTCACTCTTCCCAACAGACTCACTATATTATTCCAAAGGT 1844  
Db 1627 CATGAATTTGATTAAGTCACTCTTCCCAACAGACTCACTATATTATTCCAAAGGT 1686  
Qy 1845 GAATGCCCTGCTTATTTTCCAGCTTTGCGAAGTAGTCCAAAGAGGAGTTGAAGTGCCTAT 1904  
Db 1687 GAATGCCCTGCTTATTTTCCAGCTTTGCGAAGTAGTCCAAAGAGGAGTTGAAGTGCCTAT 1746  
Qy 1905 AGTGCAATCAATTTTGGACATGCTTCTATTGGCAACAGTGGAAACCATCCACTC 1964  
Db 1747 AGTGCAATCAATTTTGGACATGCTTCTATTGGCAACAGTGGAAACCATCCACTC 1806  
Qy 1965 ATATGACAGTGAATGAAATGCCACCACTAATTTTAACTCTCAGCTTTCAAAATACA 2024  
Db 1807 ATATGACAGTGAATGAAATGCCACCACTAATTTTAACTCTCAGCTTTCAAAATACA 1866  
Qy 2025 AACGATGAAGGTTCAAAATCAGATTAACAGTGGAGTGGACACAAAGGGAGGACCAAAA 2084  
Db 1867 AACGATGAAGGTTCAAAATCAGATTAACAGTGGAGTGGACACAAAGGGAGGACCAAAA 1926  
Qy 2085 CTGAATTTCTAGGCCAGAGGTTACGAAATTTAGTTAGTCCCATAACTCTTCTCCA 2144  
Db 1927 CTGAATTTCTAGGCCAGAGGTTACGAAATTTAGTTAGTCCCATAACTCTTCTCCA 1986  
Qy 2145 GAGCGGAAATTCCTTTTGAAGATTTCCCAAGAAACGCTTCCCGAAGTTTAAGAGA 2204  
Db 1987 GAGCGGAAATTCCTTTTGAAGATTTCCCAAGAAACGCTTCCCGAAGTTTAAGAGA 2046  
Qy 2205 CATGATGTTTAACTCAACAGGAGGCCAGGAGGATGAAATTTCCCTGGTAAATATT 2264  
Db 2047 CATGATGTTTAACTCAACAGGAGGCCAGGAGGATGAAATTTCCCTGGTAAATATT 2106  
Qy 2265 TCACCTCTTCCAAAGAGCCCGAGTTGAGTCTCAATACCTTTGGATTTTGCACCTGGACAT 2324  
Db 2107 TCACCTCTTCCAAAGAGCCCGAGTTGAGTCTCAATACCTTTGGATTTTGCACCTGGACAT 2166  
Qy 2325 GGAGACATCACTTTGAAGGATACAAATTTGTCAGAGTCAGCTTGTGAGATCATTTCTG 2384  
Db 2167 GGAGACATCACTTTGAAGGATACAAATTTGTCAGAGTCAGCTTGTGAGATCATTTCTG 2226  
Qy 2385 ATGAACATCAGCATGCTTAAATTAATAAATCAAGCTATTAATACAGATGAAACAAATGAC 2444  
Db 2227 ATGAACATCAGCATGCTTAAATTAATAAATCAAGCTATTAATACAGATGAAACAAATGAC 2286

Qy 2445 AGTTTGGTGGCTCCACAGGAAAAACAGGTTTCATAAAAGCATCTTGCCAAAACAGCTTAGGA 2504  
Db 2287 AGTTTGGTGGCTCCACAGGAAAAACAGGTTTCATAAAAGCATCTTGCCAAAACAGCTTAGGA 2346  
Qy 2505 GTGCTGAAAGATTTGACAGAGTTGATCTTTCTCGAGTGAGTGAAAGTGAATGGTCA 2564  
Db 2347 GTGCTGAAAGATTTGACAGAGTTGATCTTTCTCGAGTGAGTGAAAGTGAATGGTCA 2406  
Qy 2565 GACCAGGTCAGAAATCCACCCCTGGACTTGGAGACACAGCAAGATTTAGAGTGGAACT 2624  
Db 2407 GACCAGGTCAGAAATCCACCCCTGGACTTGGAGACACAGCAAGATTTAGAGTGGAACT 2466  
Qy 2625 CACACCCAAAAAATAGGCGGAAATGTGACAAAGAAAAAGCCCCATCTCTGATTTT 2684  
Db 2467 CACACCCAAAAAATAGGCGGAAATGTGACAAAGAAAAAGCCCCATCTCTGATTTT 2526  
Qy 2685 CCACCTGGAAGCCAGATGACAAAAAGAAAGAAATCACAGGAAAGAAAGAGAACAGT 2744  
Db 2527 CCACCTGGAAGCCAGATGACAAAAAGAAAGAAATCACAGGAAAGAAAGAGAACAGT 2586  
Qy 2745 AGAATGGAGGAAAAATGCTGAAAAATCACATAGGCGTTACTGAAAGTGTACTTTGGAAGAAAG 2804  
Db 2587 AGAATGGAGGAAAAATGCTGAAAAATCACATAGGCGTTACTGAAAGTGTACTTTGGAAGAAAG 2646  
Qy 2805 CTGCGAGCTTACACAGATAGTTACTTGGGCTTTTGGCCATGGGAGAAAAAAGTATTTT 2864  
Db 2647 CTGCGAGCTTACACAGATAGTTACTTGGGCTTTTGGCCATGGGAGAAAAAAGTATTTT 2706  
Qy 2865 CAAGATCTTCTCGACGAAGAGAGTCAATTGAAGACACAAATTTGGCATATCTTCACTGATAGC 2924  
Db 2707 CTAGATCTTCTCGACGAAGAGAGTCAATTGAAGACACAAATTTGGCATATCTTCACTGATAGC 2766  
Qy 2925 AAAATATCTGGAGGCAACTAAAGATACATTTGAGATTTCCCTCAGATATGTAAATAAA 2984  
Db 2767 AAGAATCTGGAGGCAACTAAAGATACATTTGAGATTTCCCTCAGATATGTAAATAAA 2826  
Qy 2985 ATTCTAAATAGCAAGTTTGGATTTCACTCGGGAAGTCCCTGCTCACATGCTCACATG 3044  
Db 2827 ATTCTAAATAGCAAGTTTGGATTTCACTCGGGAAGTCCCTGCTCACATGCTCACATG 2886  
Qy 3045 ATTGACCGGATTTTATGCAAGAACTGCAAGATATGTTTCCCTGGAAGATTTGCAAGAGC 3104  
Db 2887 ATTGACCGGATTTTATGCAAGAACTGCAAGATATGTTTCCCTGGAAGATTTGCAAGAGC 2946  
Qy 3105 TCATTTCAAAAGTGGCCATTTCTGAGGATATGAGTTTGGCTTCTCTTATTTTATTTAT 3164  
Db 2947 TCATTTCAAAAGTGGCCATTTCTGAGGATATGAGTTTGGCTTCTCTTATTTTATTTAT 3006  
Qy 3165 CTCATGAGTGCAGTGCAGCCACTGAAATATATCTCAAGTCTTTGATGAAAGTTGATACAGAT 3224  
Db 3007 CTCATGAGTGCAGTGCAGCCACTGAAATATATCTCAAGTCTTTGATGAAAGTTGATACAGAT 3066  
Qy 3225 CAATCTGCTGCTTCTGTCAGAGAAATCCGAACTGGCTACGAGATTCAGAACTG 3284  
Db 3067 CAATCTGCTGCTTCTGTCAGAGAAATCCGAACTGGCTACGAGATTCAGAACTG 3126  
Qy 3285 CCGTTTAAAGTTTGCAGGATTTGACAGGCTTGGAAACACATGCTTAAATAATTTGCTCAAAAAATG 3344  
Db 3127 CCGTTTAAAGTTTGCAGGATTTGACAGGCTTGGAAACACATGCTTAAATAATTTGCTCAAAAAATG 3186  
Qy 3345 CTTCTGCTGATATCACGAGCTAAATAATTTTCCCAAACTCAGGAATCTCTATCTATGAT 3404  
Db 3187 CTTCTGCTGATATCACGAGCTAAATAATTTTCCCAAACTCAGGAATCTCTATCTATGAT 3246  
Qy 3405 CCCAACCTGCCACCGCTCACTAAAGTCTAGTAACAACTGTAAACCAAGTAACTTGACAAA 3464  
Db 3247 CCCAACCTGCCACCGCTCACTAAAGTCTAGTAACAACTGTAAACCAAGTAACTTGACAAA 3306  
Qy 3465 ATCCCAAGGATTAAGGACAAAAAATAATATAGTTTGAATCATGGGAGAGAGAA 3524  
Db 3307 ATCCCAAGGATTAAGGACAAAAAATAATATAGTTTGAATCATGGGAGAGAGAA 3366



Qy	3525	ATCGCTTTTAAATGATTCGTACCAACGTTTCTCATGTGGTTGGCCAGTTGGATGACATA	3584
Db	3367	ATCGCTTTTAAATGATTCGTACCAACGTTTCTCATGTGGTTGGCCAGTTGGATGACATA	3426
Qy	3585	AGAAAAACCCCTAGGAAGTTTGTTCGCTTGAATGACAAACATTTGACCAACAATCATAAAGAT	3644
Db	3427	AGAAAAACCCCTAGGAAGTTTGTTCGCTTGAATGACAAACATTTGACCAACAATCATAAAGAT	3486
Qy	3645	GCTCAGACAGTGAAGCTGTTCTCAGGGAATTCTATGAATCATGTTCCCATACCTTCC	3704
Db	3487	GCTCAGACAGTGAAGCTGTTCTCAGGGAATTCTATGAATCATGTTCCCATACCTTCC	3546
Qy	3705	CAATTTGAACTGCCAAGAGAGTATCGAAACCGTTTCTTCATATGCATGAGCTGCAGGAA	3764
Db	3547	CAATTTGAACTGCCAAGAGAGTATCGAAACCGTTTCTTCATATGCATGAGCTGCAGGAA	3606
Qy	3765	TGAGGCGTTATCGAGACAAATTGAAGTTTTCGACCCATTGTGTACTAGCAACATTGATT	3824
Db	3607	TGAGGCGTTATCGAGACAAATTGAAGTTTTCGACCCATTGTGTACTAGCAACATTGATT	3666
Qy	3825	ATGTTTACTATATTTCTCATTTTTTGTGTGACGAGTTAAATTGCATCTAAACGGAAGATATTT	3884
Db	3667	ATGTTTACTATATTTCTCATTTTTTGTGTGACGAGTTAAATTGCATCTAAACGGAAGATATTT	3726
Qy	3885	CCAGAGAGGAGGATACACAAAGAGCTAGTCCCAATCGAATCAGAGTATAGAGATC	3941
Db	3727	CCAGAGAGGAGGATACACAAAGAGCTAGTCCCAATCGAATCAGAGTATAGAGATC	3783

RESULT 8  
 US-09-636-060C-20  
 ; Sequence 20, Application US/09636060C  
 ; Patent No. 6642038  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CANFIELD, WILLIAM M  
 ; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY  
 ; FILE REFERENCE: 210119USOCNT  
 ; CURRENT APPLICATION NUMBER: US/09/636,060C  
 ; CURRENT FILING DATE: 2000-08-10  
 ; PRIOR APPLICATION NUMBER: 60/153,831  
 ; PRIOR FILING DATE: 1999-09-14  
 ; NUMBER OF SEQ ID NOS: 52  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 20  
 ; LENGTH: 3783  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-09-636-060C-20

```
Query Match      67.2%; Score 3751; DB 4; Length 3783;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 3767; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
```

Qy	165	ATGCTGTTCAAGCTCCTGCAGAGACAAACCTATACCTGCCTGTGCCACAGGTATGGGTC	224
Db	7	ATGGGGTTCAAGCTCTTGCAGAGACAAACCTATACCTGCCTGTGCCACAGGTATGGGTC	66
Qy	225	TACGTGTGCTTCTTGGGGCGTCGTTGTACCAATCGTCTCCGCTTCCAGTTCGGAGAGGTG	284
Db	67	TACGTGTGCTTCTTGGGGCGTCGTTGTACCAATCGTCTCCGCTTCCAGTTCGGAGAGGTG	126
Qy	285	GTTCTGGAATGGAGCCGAGATCAATACCAATGTTTGGTTTGAATCCTATAGAGACAATATT	344
Db	127	GTTCTGGAATGGAGCCGAGATCAATACCATGTTTGGTTTGAATCCTATAGAGACAATATT	186
Qy	345	GCTCGAAGTCCCTTCAGAAATCGGCTTGTCTGCCCATCGGATTGACGTGTGTTTACACC	404
Db	187	GCTCGAAGAATCCTTTCAGAAATCGGCTTGTCTGCCCATCGGATTGACGTGTGTTTACACC	246
Qy	405	TGGTGAATGGCACAGATCTTGAATCTACTGAGGAACACTACGACGGTCAGAGAACAGATG	464
Db	247	TGGTGAATGGCACAGATCTTGAATCTACTGAGGAACACTACGACGGTCAGAGAACAGATG	306

Qy	465	GAGGAGGAGCAGAAAGCAATGAGAGAAATCTCTTGGGAAAAACAACAACGGAAACCTACTAAG	524
Db	307	GAGGAGGAGCAGAAAGCAATGAGAGAAATCTTGGGAAAAACAACAACGGAAACCTACTAAG	366
Qy	525	AAGAGTGAGAAAGCAGTTAGAGTGTGGCTAAACACACTGCATTAAGGTGCCAATGCTTGTA	584
Db	367	AAGAGTGAGAAAGCAGTTAGAGTGTGGCTAAACACACTGCATTAAGGTGCCAATGCTTGTC	426
Qy	585	CTGGACCCAGCCCTGCCAGCCAAACATCACCCCTGAAAGACGTGCACTCTCTTTATCTCTTCT	644
Db	427	CTGGACCCAGCCCTGCCAGCCAAACATCACCCCTGAAAGACGTGCACTCTCTTTATCTCTTCT	486
Qy	645	TTTCATTCTGCCAGTGACATTTTCAATGTGGCAAAACCAAAAGACCCCTCTCAACAATGTC	704
Db	487	TTTCATTCTGCCAGTGACATTTTCAATGTGGCAAAACCAAAAGACCCCTCTCAACAATGTC	546
Qy	705	TCAGTTGTTGTTTTTGACAGTACTAAAGATGTGGAAGATGCCCACTCTGCACTGCTTAAA	764
Db	547	TCAGTTGTTGTTTTTGACAGTACTAAAGATGTGGAAGATGCCCACTCTGCACTGCTTAAA	606
Qy	765	GGAAATAGCAGACAGACAGTATGGAGGGGTACTTGACACACAGTAAAGAGTCCCTGGGA	824
Db	607	GGAAATAGCAGACAGACAGTATGGAGGGGTACTTGACACACAGTAAAGAGTCCCTGGGA	666
Qy	825	TTAGTGTCTAATGCAAGATTTGGCTTCTCCTCAGTGGATTTCCACCAACATTTCAAGGAAACA	884
Db	667	TTAGTGTCTAATGCAAGATTTGGCTTCTCCTCAGTGGATTTCCACCAACATTTCAAGGAAACA	726
Qy	885	AATCAACTAAAAACAAAATTCGACAGAAATCTTCTCTAAAGTCAAACTGTTGCAGTTG	944
Db	727	AATCAACTAAAAACAAAATTCGACAGAAATCTTCTCTAAAGTCAAACTGTTGCAGTTG	786
Qy	945	TATTCAGAGCCAGTGTAGCGCTCTTAAACTGAATAACCCCAAGGATTTTCAAGAAATTG	1004
Db	787	TATTCAGAGCCAGTGTAGCGCTCTTAAACTGAATAACCCCAAGGATTTTCAAGAAATTG	846
Qy	1005	ANTAAGCAAACTAAGAAAGAACATGACCAATTGATGGAAAAAGAACTGACCACTAAGTCTGCA	1064
Db	847	ANTAAGCAAACTAAGAAAGAACATGACCAATTGATGGAAAAAGAACTAAGTCTGCA	906
Qy	1065	TATTTATTTGGGATCTGAGCGCCATCAGCCAGTCTTAAGCAGAGATGAAGACATCTCTGCC	1124
Db	907	TATTTATTTGGGATCTGAGCGCCATCAGCCAGTCTTAAGCAGAGATGAAGACATCTCTGCC	966
Qy	1125	AGTCGTTTTGAAGATAACGAAGAACCTGAGGTACTCTATTGCGATCTATCGAGAGGCGATGCA	1184
Db	967	AGTCGTTTTGAAGATAACGAAGAACCTGAGGTACTCTATTGCGATCTATCGAGAGGCGATGCA	1026
Qy	1185	CCATGGGTTTCGGAAATATTTTCATTTGTCACCAACGGGCAGATTCATCTCGCTGGAACCTT	1244
Db	1027	CCATGGGTTTCGGAAATATTTTCATTTGTCACCAACGGGCAGATTCATCTCGCTGGAACCTT	1086
Qy	1245	GACAATCCTCGAGTGACAAATAGTAAACACACAGAGTGTTTTTCGAAATTTGAGCCACTTG	1304
Db	1087	GACAATCCTCGAGTGACAAATAGTAAACACACAGAGTGTTTTTCGAAATTTGAGCCACTTG	1146
Qy	1305	CCTACCTTTAGTTTCACTCTGCTATTGAAAGTCAATTCATCGCATCGAAGGGCTGTCCAG	1364
Db	1147	CCTACCTTTAGTTTCACTCTGCTATTGAAAGTCAATTCATCGCATCGAAGGGCTGTCCAG	1206
Qy	1365	AAGTTTATTTTACCTTAAATGATGTCATCTTTGGGAAAGATGTCCTGCCAGATGATTTT	1424
Db	1207	AAGTTTATTTTACCTTAAATGATGTCATCTTTGGGAAAGATGTCCTGCCAGATGATTTT	1266
Qy	1425	TACAGTCACTCCAAAGGCCAGAGAGGTTTATTGACATGCGCTGTGCGCAAACCTGTGCGCAG	1484
Db	1267	TACAGTCACTCCAAAGGCCAGAGAGGTTTATTGACATGCGCTGTGCGCAAACCTGTGCGCAG	1326
Qy	1485	GGCTGCCAGGTTCTCGGATTAAGGATGGCTATTGTGACAGGCTTGTAATAATCTCAGCC	1544
Db	1327	GGCTGCCAGGTTCTCGGATTAAGGATGGCTATTGTGACAGGCTTGTAATAATCTCAGCC	1386
Qy	1545	TGCGATTTGGGATGGTGGGGATTCGCTCTGGAAACAAGTGGAGGGAGTCGTATATTGACGGA	1604

1387 TCGGATGGGATGGGGATGCTCTCGAAGACAGTGGAGGAGTGGCTATATTGACAGGA 1446  
1605 GGTGGAGGTACTGGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAAC 1664  
1447 GGTGGAGGTACTGGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAAC 1506  
1665 AGTGTCTCTTACTGTAATCAGGATGTCGGAATTCCTGGCTCGCTGATTAAGTCTGTGAC 1724  
1507 AGTGTCTCTTACTGTAATCAGGATGTCGGAATTCCTGGCTCGCTGATTAAGTCTGTGAC 1566  
1725 CAAGCATCAATGCTCTGCTGTGGGTTTGATGCTGGCGACTGTGGCGAAGATCAATTT 1784  
1567 CAAGCATCAATGCTCTGCTGTGGGTTTGATGCTGGCGACTGTGGCGAAGATCAATTT 1626  
1785 CATGAATTTGTAAGTGTATGCTCTTCTCCAAACAGACTCACTATATATTCCAAAAGGT 1844  
1627 CATGAATTTGTAAGTGTATGCTCTTCTCCAAACAGACTCACTATATATTCCAAAAGGT 1686  
1845 GAATGCTGCTCTTATTTCAGGCTTTGCAAGAGTAGCCAAAGAGGAGTTGAAGTGCCTAT 1904  
1687 GAATGCTGCTCTTATTTCAGGCTTTGCAAGAGTAGCCAAAGAGGAGTTGAAGTGCCTAT 1746  
1905 AGTGCAATCCMAATTAATTCGACATGCTTCTATTGCGCAACAGAGTGGAAACCATCCACTC 1964  
1747 AGTGCAATCCMAATTAATTCGACATGCTTCTATTGCGCAACAGAGTGGAAACCATCCACTC 1806  
1965 ATAATGACAGTGGAGTAATGCAATGCAACACCAATATATTTAATCTCACGTTTCAAAATACA 2024  
1807 ATAATGACAGTGGAGTAATGCAATGCAACACCAATATATTTAATCTCACGTTTCAAAATACA 1866  
2025 AACGATGAAGAGTTCMAAATCAGATACAGATACAGTGGAGTGGACACAAGGGAGGACCAAAA 2084  
1867 AACGATGAAGAGTTCMAAATCAGATACAGTGGAGTGGACACAAGGGAGGACCAAAA 1926  
2085 CTGAATTTCTAGGCCAGAGGGTTACGAAATTTAGTTAGTCCCATTAACACTCTTCCA 2144  
1927 CTGAATTTCTAGGCCAGAGGGTTACGAAATTTAGTTAGTCCCATTAACACTCTTCCA 1986  
2145 GAGGCGGAAATCCTTTTGGAGATATCCCAAGAAACACGCTTCCGGAAGTTTAAGAGA 2204  
1987 GAGGCGGAAATCCTTTTGGAGATATCCCAAGAAACACGCTTCCGGAAGTTTAAGAGA 2046  
2205 CATGATGTTAACTCAACAGAGAGCCCGAGGAAGAGTGAATAATTCCTCGTGAATATT 2264  
2047 CATGATGTTAACTCAACAGAGAGCCCGAGGAAGAGTGAATAATTCCTCGTGAATATT 2106  
2265 TCACCTCTTCCAAAGAGCCCGAGTGTGATCTCAATACCTTGGATTTGCAACTGGAACAT 2324  
2107 TCACCTCTTCCAAAGAGCCCGAGTGTGATCTCAATACCTTGGATTTGCAACTGGAACAT 2166  
2325 GGAGACATCACTTTGAAAGGATACAAATTTGTCGAAGTCCAGCTTGTGAGATCAATTTCTG 2384  
2167 GGAGACATCACTTTGAAAGGATACAAATTTGTCGAAGTCCAGCTTGTGAGATCAATTTCTG 2226  
2385 ATGAATCAGACATGCTTAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2444  
2227 ATGAATCAGACATGCTTAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2286  
2445 AGTTTGGTGGCTCCACAGGAAACACAGGTTTCAAAAAGCATCTTTGCCAAACAGCTTAGGA 2504  
2287 AGTTTGGTGGCTCCACAGGAAACACAGGTTTCAAAAAGCATCTTTGCCAAACAGCTTAGGA 2346  
2505 GTGTCTGAAGATTCAGAGGTGACCTTTCTCCAGTGTGATGTAAGAAGTCAATGGTCAT 2564  
2347 GTGTCTGAAGATTCAGAGGTGACCTTTCTCCAGTGTGATGTAAGAAGTCAATGGTCAT 2406  
2565 GACCAAGGTGAGAAATCCACCCCTGGACTTGGAGACCAGCAAGATTTAGAGTGGAAACT 2624  
2407 GACCAAGGTGAGAAATCCACCCCTGGACTTGGAGACCAGCAAGATTTAGAGTGGAAACT 2466  
2625 CACACCCCAAAAACCATAGGCGGAAATGTGCAAAAGAAAGCCCTCATCTCTGATTTGTT 2684

2467 CACACCCCAAAAACCATAGGCGGAAATGTGCAAAAGAAAGCCCTCATCTCTGATTTGTT 2526  
2685 CCACCTGGAAAGCCAGATGACAAAAGAAAGAAATCAAGGAAAGAAAGAAAGAAAGAAAGT 2744  
2527 CCACCTGGAAAGCCAGATGACAAAAGAAAGAAATCAAGGAAAGAAAGAAAGAAAGAAAGT 2586  
2745 AGAATCGAGGAAATGCTGAAATCAGATAGCGTTACTGAAGTGTACTTTGGAAGAAAG 2804  
2587 AGAATCGAGGAAATGCTGAAATCAGATAGCGTTACTGAAGTGTACTTTGGAAGAAAG 2646  
2805 CTGACGATTAACAGATAGTTACTTTGGGCTTTTGGCCATGGGAGAAAAAAGTATTTC 2864  
2647 CTGACGATTAACAGATAGTTACTTTGGGCTTTTGGCCATGGGAGAAAAAAGTATTTC 2706  
2865 CAAGATCTTCTCGAGAGAGAGGATCATTTGAAGACACAAATTTGGCATCTTCACTCATAGC 2924  
2707 CTAGATCTTCTCGAGAGAGAGGATCATTTGAAGACACAAATTTGGCATCTTCACTCATAGC 2766  
2925 AAAAAATCTGGAGGCAACTAAAGATACATTTGCAGATTCCTCAGATATGTAATAAAA 2984  
2767 AAAAAATCTGGAGGCAACTAAAGATACATTTGCAGATTCCTCAGATATGTAATAAAA 2826  
2985 ATTCTAAATAGCAAGTTTGGATTTCATCGCGGAAAGTCCCTGCTCAGATGCTCCTCACATG 3044  
2827 ATTCTAAATAGCAAGTTTGGATTTCATCGCGGAAAGTCCCTGCTCAGATGCTCCTCACATG 2886  
3045 ATTGACCGGATTTGTTATGCAAGAACTGCAAGATATGTTCCCTGAAAGATTTGACAAGAGC 3104  
2887 ATTGACCGGATTTGTTATGCAAGAACTGCAAGATATGTTCCCTGAAAGATTTGACAAGAGC 2946  
3105 TCATTTCAAAAGTGGCCATTTCTGAGGATATGCAAGTTTGGCTTCTCTATTTTATTAT 3164  
2947 TCATTTCAAAAGTGGCCATTTCTGAGGATATGCAAGTTTGGCTTCTCTATTTTATTAT 3006  
3165 CTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTTTGATGAAAGTTGATACAGAT 3224  
3007 CTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTTTGATGAAAGTTGATACAGAT 3066  
3225 CAATCTGGTGTCTTGTGACAGAGAAATCCGAAACACTGGCTACAGAAATTCAGAACTG 3284  
3067 CAATCTGGTGTCTTGTGACAGAGAAATCCGAAACACTGGCTACAGAAATTCAGAACTG 3126  
3285 CGGTAAAGTTTCAGAGATTTGACAGTCTGGACACATGCTAAATAATTTGCTCAAAAATG 3344  
3127 CGGTAAAGTTTCAGAGATTTGACAGTCTGGACACATGCTAAATAATTTGCTCAAAAATG 3186  
3345 CTTCTCTGTGATATCAGCAGCTAAATATATTTCCCAACACTCAGGAAATCTCTACTATGAT 3404  
3187 CTTCTCTGTGATATCAGCAGCTAAATATATTTCCCAACACTCAGGAAATCTCTACTATGAT 3246  
3405 CCAACCTGCCACCGGTCACTAAAGTCTAGTAAACAACTGTAAACCACTGTAACCACTGACAAA 3464  
3247 CCAACCTGCCACCGGTCACTAAAGTCTAGTAAACAACTGTAAACCACTGTAACCACTGACAAA 3306  
3465 ATCCCAAAAGCATATAAGGACAAAAACAATATAGGTTTGAATCATGGGAGAAAGAA 3524  
3307 ATCCCAAAAGCATATAAGGACAAAAACAATATAGGTTTGAATCATGGGAGAAAGAA 3366  
3525 ATCGCTTTTAAATGATTCGTACCAAGCTTCTCATGTTGGTTGGCCAGTTGGATGACATA 3584  
3367 ATCGCTTTTAAATGATTCGTACCAAGCTTCTCATGTTGGTTGGCCAGTTGGATGACATA 3426  
3585 AGAAAAACCCCTAGCAAGTTTGTTCCTGTAATGACAACTTGACCACTCATATAAAGAT 3644  
3427 AGAAAAACCCCTAGCAAGTTTGTTCCTGTAATGACAACTTGACCACTCATATAAAGAT 3486  
3645 GGTGACAGAGTGAAGGCTTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCC 3704  
3487 GGTGACAGAGTGAAGGCTTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCC 3546  
3705 CAATTTGAACCTGCCAAGAGATATCGAAACGTTTCTTCTATATGATGACCTGACAGAA 3764  
3547 CAATTTGAACCTGCCAAGAGATATCGAAACGTTTCTTCTATATGATGACCTGACAGAA 3606

QY 3765 TGGAGGGCTTATCGAGACAAATTTGAAGTTTGGAGCCCATTTGTACTAGCAACATTTGATT 3824  
Db |||||  
QY 3607 TGGAGGGCTTATCGAGACAAATTTGAAGTTTGGAGCCCATTTGTACTAGCAACATTTGATT 3666  
Db |||||  
QY 3825 ATGTTTACTATATTTCTCATTTTTTTTGTCTGAGCAGTTAAATTGCATTTAAGCGGAAGATATTT 3884  
Db |||||  
QY 3667 ATGTTTACTATATTTCTCATTTTTTTTGTCTGAGCAGTTAAATTGCATTTAAGCGGAAGATATTT 3726  
Db |||||  
QY 3885 CCCAGAGGAGATACACAAAGAGCTAGTCCCATCGAATCAGATATAGAGATC 3941  
Db |||||  
QY 3727 CCCAGAGGAGATACACAAAGAGCTAGTCCCATCGAATCAGATATAGAGATC 3783  
Db |||||

RESULT 9

US-09-986-552-20  
; Sequence 20, Application US/09986552  
; Patent No. 6670165  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 215089US77DIV  
; CURRENT APPLICATION NUMBER: US/09/986,552  
; CURRENT FILING DATE: 2001-11-09  
; PRIOR APPLICATION NUMBER: 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 20  
; LENGTH: 3783  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-986-552-20

Query Match 67.2%; Score 3761; DB 4; Length 3783;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 3767; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
QY 165 ATGCTGTCTCAAGCTCTCTGAGAGACAAACCTATACCTGCTGCTCCACAGGTATGGGCTC 224  
Db 7 ATGGGGTTCAAGCTCTTGCAGAGACAAACCTATACCTGCTGCTCCACAGGTATGGGCTC 66  
QY 225 TAGCTGTGCTTTTGGGCGCTGTTGTGACCATCGTCTCGCCCTTCCAGTTCGGAGAGTG 284  
Db 67 TAGCTGTGCTTTTGGGCGCTGTTGTGACCATCGTCTCGCCCTTCCAGTTCGGAGAGTG 126  
QY 285 GTTCTGGAATGGAGCGGAGATCAATACCATGTTTGTGATTTGTTGATTCCTATAGACAAATTT 344  
Db 127 GTTCTGGAATGGAGCGGAGATCAATACCATGTTTGTGATTTGTTGATTCCTATAGACAAATTT 186  
QY 345 GCTGGAAGTCCCTTTTCAAGATCGGCTTTGCTGCCCCATGCCGATTCGAGCTTGTGTTTACACC 404  
Db 187 GCTGGAAGTCCCTTTTCAAGATCGGCTTTGCTGCCCCATGCCGATTCGAGCTTGTGTTTACACC 246  
QY 405 TGGGTGAATGGACAGATCTTGAATCTAGAGGAACTACAGCAGGTACAGAGACAGATG 464  
Db 247 TGGGTGAATGGACAGATCTTGAATCTAGAGGAACTACAGCAGGTACAGAGACAGATG 306  
QY 465 GAGGAGGAGCAGAAAGCAATGAGAGAAATCCCTTGGGAAAAACACAAACGAACTACTAAG 524  
Db 307 GAGGAGGAGCAGAAAGCAATGAGAGAAATCCCTTGGGAAAAACACAAACGAACTACTAAG 366  
QY 525 AAGAGTGAGAGCAGTTAGAGTTTGGCTTAACACTGCAATTAAGTGGCAATGCTTGTGA 584  
Db 367 AAGAGTGAGAGCAGTTAGAGTTTGGCTTAACACTGCAATTAAGTGGCAATGCTTGTGC 426  
QY 585 CTGGACCCAGCCCTGCGCAGCAACATCACCTGAGAGAGTCCCTCTCTTTATCTCTCT 644  
Db 427 CTGGACCCAGCCCTGCGCAGCAACATCACCTGAGAGAGTCCCTCTCTTTATCTCTCT 486  
QY 645 TTTCAATTCGCCAGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGTC 704

Db |||||  
QY 487 TTTCAATTCGCCAGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGTC 546  
QY 705 TCAGTTGTTGTTTTCAGCAGTACTAAGAGTGTGGAAGTGCCTCCTCTGACATGCTTAAA 764  
Db |||||  
QY 547 TCAGTTGTTGTTTTCAGCAGTACTAAGAGTGTGGAAGTGCCTCCTCTGACATGCTTAAA 606  
Db |||||  
QY 765 GGAATAAGCAGACAGCAGTATGAGGGGGTACTTGCACAAACAGATATAAAGAGTCCCTGGA 824  
Db |||||  
QY 607 GGAATAAGCAGACAGCAGTATGAGGGGGTACTTGCACAAACAGATATAAAGAGTCCCTGGA 666  
QY 825 TTAGTGCTAATCAAGATTTGGCTTTCTGAGTGGATTTCCACCAACATTTCAAGGAAACA 884  
Db |||||  
QY 667 TTAGTGCTAATCAAGATTTGGCTTTCTGAGTGGATTTCCACCAACATTTCAAGGAAACA 726  
QY 885 AATCAACTAAACAAATTTGCCAGAAAATCTTTCTCTAAAGTCAAACTGTTGCGAGTTG 944  
Db |||||  
QY 727 AATCAACTAAACAAATTTGCCAGAAAATCTTTCTCTAAAGTCAAACTGTTGCGAGTTG 786  
QY 945 TATTTCAGAGCCAGTGTAGCGCTTCTAAACTGAATAACCCCAAGGATTTTCAAGAAATTG 1004  
Db |||||  
QY 787 TATTTCAGAGCCAGTGTAGCGCTTCTAAACTGAATAACCCCAAGGATTTTCAAGAAATTG 846  
QY 1005 AATAAGCAAACTAAGAGAACATGACCATTTGATGGAAGAAAGTGCACCAATGCTGCGCA 1064  
Db |||||  
QY 847 AATAAGCAAACTAAGAGAACATGACCATTTGATGGAAGAAAGTGCACCAATGCTGCGCA 906  
QY 1065 TATTATTATGAGGATCTGAGCGGCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCC 1124  
Db |||||  
QY 907 TATTATTATGAGGATCTGAGCGGCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCC 966  
QY 1125 AGTCGTTTTGAAGATAACGAAGAACTGAGGTACTCATTTGGGATCTATCGAGAGGCGATGCA 1184  
Db |||||  
QY 967 AGTCGTTTTGAAGATAACGAAGAACTGAGGTACTCATTTGGGATCTATCGAGAGGCGATGCA 1026  
QY 1185 CCATGGGTTCCGGAATATTTTTCATTGTCAACCAACGGGCGAGATTCATCTCGCTGGAACCTT 1244  
Db |||||  
QY 1027 CCATGGGTTCCGGAATATTTTTCATTGTCAACCAACGGGCGAGATTCATCTCGCTGGAACCTT 1086  
QY 1245 GACAATCCTCGAGTGACAATAGTAACACACACAGATGTTTTTTCGAATTTGAGGCCACTTG 1304  
Db |||||  
QY 1087 GACAATCCTCGAGTGACAATAGTAACACACACAGATGTTTTTTCGAATTTGAGGCCACTTG 1146  
QY 1305 CCTACCTTTAGTTTCACTGCTATTGAAAGTCACATTCATCGCATCGAAGGGCTGTCACG 1364  
Db |||||  
QY 1147 CCTACCTTTAGTTTCACTGCTATTGAAAGTCACATTCATCGCATCGAAGGGCTGTCACG 1206  
QY 1365 AAGTTTATTTTACCTAAATGATGATGTCATGTTTGGGAAGGATGCTGCGCCAGATGATTTT 1424  
Db |||||  
QY 1207 AAGTTTATTTTACCTAAATGATGATGTCATGTTTGGGAAGGATGCTGCGCCAGATGATTTT 1266  
QY 1425 TACAGTCACTCCAAAGGCCAGAGGTTTATTTGACATGGCTGTGCCAAATCTGTGCCGAG 1484  
Db |||||  
QY 1267 TACAGTCACTCCAAAGGCCAGAGGTTTATTTGACATGGCTGTGCCAAATCTGTGCCGAG 1326  
QY 1485 GGCTGCCAGGTTCTCGATTAAAGGATGGCTATTGTGACAAAGGCTTGTATAATTTCAGCC 1544  
Db |||||  
QY 1327 GGCTGCCAGGTTCTCGATTAAAGGATGGCTATTGTGACAAAGGCTTGTATAATTTCAGCC 1386  
QY 1545 TGCATTTGGATGAGTGGGATTTGCTCTGGAACAAGTGGAGGAGTCCCTATATTGCAAGGA 1604  
Db |||||  
QY 1387 TGCATTTGGATGAGTGGGATTTGCTCTGGAACAAGTGGAGGAGTCCCTATATTGCAAGGA 1446  
QY 1605 GGTGAGGTAAGTGGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGTGGAGGAATAAAC 1664  
Db |||||  
QY 1447 GGTGAGGTAAGTGGGAGTATTGGAGTTGGAGTTGGACACCCCTGGCAGTTTGTGGAGGAATAAAC 1506  
QY 1665 AGTGTCTTCTTACTGTAAATCAGGGATGTCGAATTCCTGGCTCGCTGATAGTGTCTGAGC 1724  
Db |||||  
QY 1507 AGTGTCTTCTTACTGTAAATCAGGGATGTCGAATTCCTGGCTCGCTGATAGTGTCTGAGC 1566  
QY 1725 CAAGCATGCAATGCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1784



RESULT 10

```

US-09-636-596C-20
; Sequence 20, Application US/09636596C
; Patent No. 6770468
; GENERAL INFORMATION:
; APPLICANT: CAMFIELD, WILLIAM
; TITLE OF INVENTION: PHOSPHODIESTER A
; FILE REFERENCE: 10929-0001-77
; CURRENT APPLICATION NUMBER: US/09/636
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 20
; LENGTH: 3783
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-636-596C-20

```

```
Query Match          67.2%; Score 3761; DB 4; Length 3783;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 3767; Conservative 0; Mismatches 10; Indels 0; Gaps 0
```

165	ATGCTGTTCAAGCTCTCGACAGACAAACCTATACCTGCGCTGTCCACACAGGTATGGCTC	221	TTAGTCTGAATGAGCGGAGATCAATACCATGTTTGTTCCTATAGACACAATATT	344
Qy				
7	ATGGGGTTCAAGCTCTTGACAGACAAACCTATACCTGCGCTGTCCACACAGGTATGGGCTC	66	GTTCCTGGAATGAGCGCGAGATCAATACCATGTTTGTTCCTATAGACACAATATT	186
Db				
225	TACGCTGTGCTTCTTGGGGGTGGTGTGTCACCATCGTCTCCGCGCTTCAGATTCGGAGAGGTG	284	GTTCCTGGAATGAGCGCGAGATCAATACCATGTTTGTTCCTATAGACACAATATT	344
Qy				
67	TACGCTGTGCTTCTTGGGGGTGGTGTGTCACCATCGTCTCCGCGCTTCAGATTCGGAGAGGTG	126	GTTCCTGGAATGAGCGCGAGATCAATACCATGTTTGTTCCTATAGACACAATATT	186
Db				
345	GCTGGAAGTCCCTTTCAGAACTCGGCTTGTCTGCCCATGCCGATTGACGTTGTTTACACC	404	GTTCCTGGAATGAGCGCGAGATCAATACCATGTTTGTTCCTATAGACACAATATT	344
Qy				
187	GCTGGAAGTCCCTTTCAGAACTCGGCTTGTCTGCCCATGCCGATTGACGTTGTTTACACC	246	GTTCCTGGAATGAGCGCGAGATCAATACCATGTTTGTTCCTATAGACACAATATT	186
Db				
405	TGGGTGAATGGCACAGACTCTTGAACCTACTGAAGGAACTACAGCAGGTCAGAGAACAGATG	464	GTTCCTGGAATGAGCGCGAGATCAATACCATGTTTGTTCCTATAGACACAATATT	344
Qy				
247	TGGGTGAATGGGCACAGACTCTTGAACCTACTGAAGGAACTACAGCAGGTCAGAGAACAGATG	306	GTTCCTGGAATGAGCGCGAGATCAATACCATGTTTGTTCCTATAGACACAATATT	186
Db				
465	GAGGGGACGAGAAAGCAATGAGAGAAATCCTTCGSGAAAACACAAACGGAACCTACTAAG	524	GTTCCTGGAATGAGCGCGAGATCAATACCATGTTTGTTCCTATAGACACAATATT	344
Qy				
307	GAGGAGGAGCAGAAAGCAATGAGAGAAATCCTTCGSGAAAACACAAACGGAACCTACTAAG	366	GTTCCTGGAATGAGCGCGAGATCAATACCATGTTTGTTCCTATAGACACAATATT	344
Db				
525	AAGAGTGAGAAACGAGTTAGAGTGTTTGCTAACAACCTGTCATTAAAGGTGCCAATGCTTGA	584	GTTCCTGGAATGAGCGCGAGATCAATACCATGTTTGTTCCTATAGACACAATATT	344
Qy				
367	AAGAGTGAGAAACGAGTTAGAGTGTTTGCTAACAACCTGTCATTAAAGGTGCCAATGCTTGA	426	GTTCCTGGAATGAGCGCGAGATCAATACCATGTTTGTTCCTATAGACACAATATT	344
Db				
585	CTGGACCCAGCCCTGCCACCAACATACCCCTGAGGACGTCGCATCTCTTTATCCCTCT	644	GTTCCTGGAATGAGCGCGAGATCAATACCATGTTTGTTCCTATAGACACAATATT	344
Qy				
427	CTGGACCCAGCCCTGCCACCAACATACCCCTGAGGACGTCGCATCTCTTTATCCCTCT	486	GTTCCTGGAATGAGCGCGAGATCAATACCATGTTTGTTCCTATAGACACAATATT	344
Db				
645	TTTCACTCTGCCAGTGACATTTTCAATGTGCAAAACCAAAAACCCCTTACCAATATGC	704	GTTCCTGGAATGAGCGCGAGATCAATACCATGTTTGTTCCTATAGACACAATATT	344
Qy				
487	TTTCACTCTGCCAGTGACATTTTCAATGTGCAAAACCAAAAACCCCTTACCAATATGC	546	GTTCCTGGAATGAGCGCGAGATCAATACCATGTTTGTTCCTATAGACACAATATT	344
Db				
705	TCAGTTGTTGTTTTTGAACGACTACTAAGGATGTTGGAAGATGCCCACTCTGACCTGCTTAA	764	GTTCCTGGAATGAGCGCGAGATCAATACCATGTTTGTTCCTATAGACACAATATT	344
Qy				
547	TCAGTTGTTGTTTTTGAACGACTACTAAGGATGTTGGAAGATGCCCACTCTGACCTGCTTAA	606	GTTCCTGGAATGAGCGCGAGATCAATACCATGTTTGTTCCTATAGACACAATATT	344
Db				
765	GGAAATAGCAGACAGACGATGAGAGGGGTACTTTGACACAGATAAAGAAAGTCCCTGGA	824	GTTCCTGGAATGAGCGCGAGATCAATACCATGTTTGTTCCTATAGACACAATATT	344
Qy				
607	GGAAATAGCAGACAGACGATGAGAGGGGTACTTTGACACAGATAAAGAAAGTCCCTGGA	666	GTTCCTGGAATGAGCGCGAGATCAATACCATGTTTGTTCCTATAGACACAATATT	344
Db				
825	TTAGTCTGAATGCAAGATTGCGCTTTCCTGAGTGGATTTCACCAACATTCAGAGGAAACA	884	GTTCCTGGAATGAGCGCGAGATCAATACCATGTTTGTTCCTATAGACACAATATT	344
Qy				
667	TTAGTCTGAATGCAAGATTGCGCTTTCCTGAGTGGATTTCACCAACATTCAGAGGAAACA	726	GTTCCTGGAATGAGCGCGAGATCAATACCATGTTTGTTCCTATAGACACAATATT	344
Db				

Qy	885	AATCACTAAAA	CAAAAT	TGCAGAAATCTTTCTCTTAAAGCTCAAACTGTTGCAAGTTG	944
Db	727	AATCACTAAAA	CAAAAT	TGCAGAAATCTTTCTCTTAAAGCTCAAACTGTTGCAAGTTG	786
Qy	945	TATTCAGAGCC	AGTGTAGCGCTTCTAAAACTCAATAAACCCCAAGGATTTTCAAGAAATTG	1004	
Db	787	TATTCAGAGCC	AGTGTAGCGCTTCTAAAACTCAATAAACCCCAAGGATTTTCAAGAAATTG	846	
Qy	1005	AATAAGCAAACTA	AGAGAAACATGACCATGTATGAGAAAGAAATCTGACCATTAAGTCTCTGCA	1064	
Db	847	AATAAGCAAACTA	AGAGAAACATGACCATGTATGAGAAAGAAATCTGACCATTAAGTCTCTGCA	906	
Qy	1065	TATTTATATTGGAT	CTGAGCGCCATCAGCCAGCTTAAGCAGGATCAAGACATCTCTGCC	1124	
Db	907	TATTTATATTGGAT	CTGAGCGCCATCAGCCAGCTTAAGCAGGATCAAGACATCTCTGCC	966	
Qy	1125	AGTCGTTTTGAAGAT	AAACGAAGAACTAGAGTACTCAATTTGGATCTATPCGAGAGGCGATGCA	1184	
Db	967	AGTCGTTTTGAAGAT	AAACGAAGAACTAGAGTACTCAATTTGGATCTATPCGAGAGGCGATGCA	1026	
Qy	1185	CCATGGGTTCCGGAAT	ATTTTTCATTTGTCCAAACGGCGAGATTCATCCTCGCTGAACCTT	1244	
Db	1027	CCATGGGTTCCGGAAT	ATTTTTCATTTGTCCAAACGGCGAGATTCATCCTCGCTGAACCTT	1086	
Qy	1245	GACAACTCCTCGAGT	GCACAATAGTAACACACCCAGGATTTTTTTCGAAATTTTGAGCCACTTG	1304	
Db	1087	GACAACTCCTCGAGT	GCACAATAGTAACACACCCAGGATTTTTTTCGAAATTTTGAGCCACTTG	1146	
Qy	1305	CCTACCTTTAGTTCA	CCCTGCTATTGAAAGTCAATTCATCGCATCGAAGGGCTGTCCAG	1364	
Db	1147	CCTACCTTTAGTTCA	CCCTGCTATTGAAAGTCAATTCATCGCATCGAAGGGCTGTCCAG	1206	
Qy	1365	AAGTTTTATTACCT	TAATATGATGTATCATCTTTGGGAAGGATGCTCGCCAGATGATTTT	1424	
Db	1207	AAGTTTTATTACCT	TAATATGATGTATCATCTTTGGGAAGGATGCTCGCCAGATGATTTT	1266	
Qy	1425	TACAGTCACTCC	AAAAGGCCAAGAGTTTATTTGACATGGCTGTGCCAAACTGTGCCGAG	1484	
Db	1267	TACAGTCACTCC	AAAAGGCCAAGAGTTTATTTGACATGGCTGTGCCAAACTGTGCCGAG	1326	
Qy	1485	GGCTGCCCCAGGTT	CTCGATTAAGGATGGCTATTGTGCAAGGCTCTGTAATAATTCAAGCC	1544	
Db	1327	GGCTGCCCCAGGTT	CTCGATTAAGGATGGCTATTGTGCAAGGCTCTGTAATAATTCAAGCC	1386	
Qy	1545	TGCGATTTGGAT	GGTGGGATCTCTCTGGAACAGTGAAGGAGTGCCTATATTTGCAAGGA	1604	
Db	1387	TGCGATTTGGAT	GGTGGGATCTCTCTGGAACAGTGAAGGAGTGCCTATATTTGCAAGGA	1446	
Qy	1605	GGTGGAGGTACT	GGGAGTATTGAGTTTGGACACCCCTCGGCAGTTTGGTGGAGGAATAAAC	1664	
Db	1447	GGTGGAGGTACT	GGGAGTATTGAGTTTGGACACCCCTCGGCAGTTTGGTGGAGGAATAAAC	1506	
Qy	1665	AGTGTCTCTTACT	GTATAACAGGATGTGCGAAATTCCTGGCTCGGTGATAAGTTCTGTGAC	1724	
Db	1507	AGTGTCTCTTACT	GTATAACAGGATGTGCGAAATTCCTGGCTCGGTGATAAGTTCTGTGAC	1566	
Qy	1725	CAAGCATGCAAT	GTCTTGTGCTGGTTGATGTGGCGACTGTGGCGAAGATCATTTT	1784	
Db	1567	CAAGCATGCAAT	GTCTTGTGCTGGTTGATGTGGCGACTGTGGCGAAGATCATTTT	1626	
Qy	1785	CATGAATTGTATA	AGTGTATCTTCTCCAAACACAGACTCACTATATTTTCCAAAAGGT	1844	
Db	1627	CATGAATTGTATA	AGTGTATCTTCTCCAAACACAGACTCACTATATTTTCCAAAAGGT	1686	
Qy	1845	GAATGCTGCTCT	TATTTTTCAGCTTTGCAAGATGAGTGCCTTAAAGAGGTTGAAGTGCCTAT	1904	
Db	1687	GAATGCTGCTCT	TATTTTTCAGCTTTGCAAGATGAGTGCCTTAAAGAGGTTGAAGTGCCTAT	1746	
Qy	1905	AGTGACAAATCCA	TAAATTCGACATGCTTCTTATTTGCCAACAGTCGAAAAACCATCCACCTC	1964	
Db	1747	AGTGACAAATCCA	TAAATTCGACATGCTTCTTATTTGCCAACAGTCGAAAAACCATCCACCTC	1806	
Qy	1965	ATAATGCACAGT	GGAAATGAATGCCAACCAATACATTTTAACTCAAGTTTCAAAATPACA	2024	







1



Db 415 CTGCAGCAGCAATCACCTCGAGGACCTGCACATCTCTTTATCCTCTTTTCAATTCGCC 474  
Qy 657 AGTGACATTTCAATGTTGCGAAACCAAAACCCCTCTACCAATGCTCAGTTGTTGTT 716  
Db 475 AGTGACATTTCAATGTTGCGAAACCAAAACCCCTCTACCAATGCTCAGTTGTTGTT 534  
Qy 717 TTGACAGTACTAAGAGTGTGAAGATGCCCACTCTGACATGCTTTAAAGGAAATAGCAGA 776  
Db 535 TTGACAGTACTAAGAGTGTGAAGATGCCCACTCTGACATGCTTTAAAGGAAATAGCAGA 594  
Qy 777 CAGACAGTATGGAGGGGTACTTGACAACAGATAAAGAAAGTCCCTGGATAGTCTAATG 836  
Db 595 CAGACAGTATGGAGGGGTACTTGACAACAGATAAAGAAAGTCCCTGGATAGTCTAATG 654  
Qy 837 CAAGATTTGGCTTCTCAGTGGATTTCCACCAATTTCAAGGAAACAAATCAACTAAAA 896  
Db 655 CAAGATTTGGCTTCTCAGTGGATTTCCACCAATTTCAAGGAAACAAATCAACTAAAA 714  
Qy 897 ACAAAATTCGCCAGAAAATCTTTCTCTAAAGTCAAACTGTTGCAAGTCTGTTTCAAGGCC 956  
Db 715 ACAAAATTCGCCAGAAAATCTTTCTCTAAAGTCAAACTGTTGCAAGTCTGTTTCAAGGCC 774  
Qy 957 AGTGTAGGCTTCTAAACTGATTAACCCCAAGGATTTTCAAGATTTGAATTAAGCAACT 1016  
Db 775 AGTGTAGGCTTCTAAACTGATTAACCCCAAGGATTTTCAAGATTTGAATTAAGCAACT 834  
Qy 1017 AAGAGAAATGACCAATGATGGAAGAAAGTCAACATAAGTCTCTGCAATTTTAAATG 1076  
Db 835 AAGAGAAATGACCAATGATGGAAGAAAGTCAACATAAGTCTCTGCAATTTTAAATG 894  
Qy 1077 GATCTGAGCGCATCAGCCAGTCTAAGCAGATGAAGACATCTCTGCCAGTCTGTTTGA 1136  
Db 895 GATCTGAGCGCATCAGCCAGTCTAAGCAGATGAAGACATCTCTGCCAGTCTGTTTGA 954  
Qy 1137 GATACGAAGAACTCAGGTACTCATTTGGATCTATCGAGAGGATGCAACCTGTTGCG 1196  
Db 955 GATACGAAGAACTCAGGTACTCATTTGGATCTATCGAGAGGATGCAACCTGTTGCG 1014  
Qy 1197 AATAATTTTCAATGTCACCAAGGAGATTCATCTCGCTGAACTTTGACAACTCTCGA 1256  
Db 1015 AATAATTTTCAATGTCACCAAGGAGATTCATCTCGCTGAACTTTGACAACTCTCGA 1074  
Qy 1257 GTGACAAATGATTAACACCAAGATGTTTTCGAAATTTGAGCCACTTCCCTACCTTAGT 1316  
Db 1075 GTGACAAATGATTAACACCAAGATGTTTTCGAAATTTGAGCCACTTCCCTACCTTAGT 1134  
Qy 1317 TCACCTGCTATTGAAAGTCAATTCATCGCATCGAAGGCTGTCCGAGAGTATTATTAC 1376  
Db 1135 TCACCTGCTATTGAAAGTCAATTCATCGCATCGAAGGCTGTCCGAGAGTATTATTAC 1194  
Qy 1377 CTAAATGATGATGTCATGTTGGGAAAGATGTCGGCAGATGATTTTACAGTCACTCC 1436  
Db 1195 CTAAATGATGATGTCATGTTGGGAAAGATGTCGGCAGATGATTTTACAGTCACTCC 1254  
Qy 1437 AAGGCCAGAGGTTTATTGACATGGCTGTGCCAAACTGTGCCAGGGCTGCCAGGT 1496  
Db 1255 AAGGCCAGAGGTTTATTGACATGGCTGTGCCAAACTGTGCCAGGGCTGCCAGGT 1314  
Qy 1497 TCCTGGATTAAGATGGCTTATTGACAAAGGCTGTGAATAATTCAGCCTGCGATTGGAT 1556  
Db 1315 TCCTGGATTAAGATGGCTTATTGACAAAGGCTGTGAATAATTCAGCCTGCGATTGGAT 1374  
Qy 1557 GGTGGGATGCTCTGGAACAGTGGAGGAGTCTCTATATTGCGAGGAGTGGAGGTACT 1616  
Db 1375 GGTGGGATGCTCTGGAACAGTGGAGGAGTCTCTATATTGCGAGGAGTGGAGGTACT 1434  
Qy 1617 GGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGTGGAGGAATAAACAGTGTCTCTTAC 1676  
Db 1435 GGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGTGGAGGAATAAACAGTGTCTCTTAC 1494  
Qy 1677 TGTAAATCAGGATGTGCGAATTCCTGGCTCGCTGATTAAGTTCTGTGACCAAGCATGCAAT 1736

Db 1495 TGTAAATCAGGATGTGCGAATTCCTGGCTCGCTGATTAAGTTCTGTGACCAAGCATGCAAT 1554  
Qy 1737 GTCTTCTCTGTGGGTTTGTGATGCTGCGACTGTGGSCAAGATCATTTTTCATGAATGTAT 1796  
Db 1555 GTCTTCTCTGTGGGTTTGTGATGCTGCGACTGTGGSCAAGATCATTTTTCATGAATGTAT 1614  
Qy 1797 AAAGTGTATCTTCTCTCCAAACAGACTCTATATATTATTTCCAAAGGTGAATGCTGCT 1856  
Db 1615 AAAGTGTATCTTCTCTCCAAACAGACTCTATATATTATTTCCAAAGGTGAATGCTGCT 1674  
Qy 1857 TATTTTCAGCTTTGCGAAGTAGCCAAAGAGAGTGAAGTGCCTATAGTCACATCCA 1916  
Db 1675 TATTTTCAGCTTTGCGAAGTAGCCAAAGAGAGTGAAGTGCCTATAGTCACATCCA 1734  
Qy 1917 ATAATTCGACATGCTTCTTATTTGCCAAACAGTGGAAACCATCCACCTCTATAATGCACT 1976  
Db 1735 ATAATTCGACATGCTTCTTATTTGCCAAACAGTGGAAACCATCCACCTCTATAATGCACT 1794  
Qy 1977 GGAATGATGCTCCACCAATATATTTTATCTCACTTTTCAAAATACAAACGATGAAG 2036  
Db 1795 GGAATGATGCTCCACCAATATATTTTATCTCACTTTTCAAAATACAAACGATGAAG 1854  
Qy 2037 TTCAAATGCGAGATAACAGTGGAGGTGGACACAAGGGAGGACCAAACTGTAATTTCTACG 2096  
Db 1855 TTCAAATGCGAGATAACAGTGGAGGTGGACACAAGGGAGGACCAAACTGTAATTTCTACG 1914  
Qy 2097 GCCCAGAGGGTTACGAAAATTTAGTTAGTCCCATTAACCTTTTCCAGAGGCGGAAATC 2156  
Db 1915 GCCCAGAGGGTTACGAAAATTTAGTTAGTCCCATTAACCTTTTCCAGAGGCGGAAATC 1974  
Qy 2157 CTTTTCAGGATATTTCCCAAGAAACCGCTTCCCAAGTTTAAAGAGACATGATGTTAAC 2216  
Db 1975 CTTTTCAGGATATTTCCCAAGAAACCGCTTCCCAAGTTTAAAGAGACATGATGTTAAC 2034  
Qy 2217 TCACCAAGGAGAGCCAGGAAAGAGTGAATAATTTCCCTCTGGTAAATTTTCACTCTCTCCA 2276  
Db 2035 TCACCAAGGAGAGCCAGGAAAGAGTGAATAATTTCCCTCTGGTAAATTTTCACTCTCTCCA 2094  
Qy 2277 AAAGAGCCCGAGTGTAGTCTCAATACCTTTGGATTTGCAACTGGAAATGAGACATGACT 2336  
Db 2095 AAAGAGCCCGAGTGTAGTCTCAATACCTTTGGATTTGCAACTGGAAATGAGACATGACT 2154  
Qy 2337 TTCAAGAGATACAAATTTGTCAGAGTCAGCTTGTCTGAGATCATTTCTGATGAATCACAG 2396  
Db 2155 TTCAAGAGATACAAATTTGTCAGAGTCAGCTTGTCTGAGATCATTTCTGATGAATCACAG 2214  
Qy 2397 CATGCTAAATAAAAAATCAAGCTATAATAACAGATGAATAACAAATGACAGTTTGGTGGCT 2456  
Db 2215 CATGCTAAATAAAAAATCAAGCTATAATAACAGATGAATAACAAATGACAGTTTGGTGGCT 2274  
Qy 2457 CCACAGGAAACAGGTTTCATTAAGCATCTTCCCAACAGCTTTAGAGTGTCTGAAAGA 2516  
Db 2275 CCACAGGAAACAGGTTTCATTAAGCATCTTCCCAACAGCTTTAGAGTGTCTGAAAGA 2334  
Qy 2517 TTGACAGGTTGACTTTTCTGCAAGTGAATGTAATAAGTGAATGCTCATGCCAGGTCAG 2576  
Db 2335 TTGACAGGTTGACTTTTCTGCAAGTGAATGTAATAAGTGAATGCTCATGCCAGGTCAG 2394  
Qy 2577 AATCCACCCCTGAGCTTGGAGACACAGCAGAGATTTAGAGTGAATACTCACACCAAAA 2636  
Db 2395 AATCCACCCCTGAGCTTGGAGACACAGCAGAGATTTAGAGTGAATACTCACACCAAAA 2454  
Qy 2637 ACCATAGCCGGAATGTGACAAAGAAAGCCCCCATCTCTGATTTTCCACTGGAAAGC 2696  
Db 2455 ACCATAGCCGGAATGTGACAAAGAAAGCCCCCATCTCTGATTTTCCACTGGAAAGC 2514  
Qy 2697 CAGATGACAAAGAAAGAAATACAGGGAAGAAAGAAAGAGACAGTAGATGAGGAA 2756  
Db 2515 CAGATGACAAAGAAAGAAATACAGGGAAGAAAGAAAGAGACAGTAGATGAGGAA 2574  
Qy 2757 AATGCTGAAATCACATAGGCTTACTGAAAGTGTACTTGGAAAGAAAGCTGACAGCTTAC 2816  
Db 2575 AATGCTGAAATCACATAGGCTTACTGAAAGTGTACTTGGAAAGAAAGCTGACAGCTTAC 2634

2817 ACAGATAGTTACTTGGGCTTTTGGCATGGGAGAAAAAAGATATTTCCAGATCTTCTC 2876  
2635 ACAGATAGTTACTTGGGCTTTTGGCATGGGAGAAAAAAGATATTTCTAGATCTTCTC 2694  
2877 GACGAGAGAGTCTATTCAGACACAAATGGCATCTTCTACTGATAGCAAAAAATCTCTGG 2936  
2695 GACGAGAGAGTCTATTCAGACACAAATGGCATCTTCTACTGATAGCAAAAAATCTCTGG 2754  
2937 AGGCAACTAAAGATACATTTGCGAATTCCTCGATATGTAATAAATTTCTAAATAGC 2996  
2755 AGGCAACTAAAGATACATTTGCGAATTCCTCGATATGTAATAAATTTCTAAATAGC 2814  
2997 AGTTTGGATTACATTCGGGAAAGTCCCTGCTCACATGCTCTCATGATTCGCGGAT 3056  
2815 AGTTTGGATTACATTCGGGAAAGTCCCTGCTCACATGCTCTCATGATTCGCGGAT 2874  
3057 GTTATGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAAGAGCTCATTTCAAAA 3116  
2875 GTTATGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAAGAGCTCATTTCAAAA 2934  
3117 GTGCGCCATTTGAGGATATGAGTTTGGCTTCTCTTATTTTATATCTCATGAGTGCA 3176  
2935 GTGCGCCATTTGAGGATATGAGTTTGGCTTCTCTTATTTTATATCTCATGAGTGCA 2994  
3177 GTGCGCCATTTGAGGATATGAGTTTGGCTTCTCTTATTTTATATCTCATGAGTGCA 3236  
2995 GTGCGCCATTTGAGGATATGAGTTTGGCTTCTCTTATTTTATATCTCATGAGTGCA 3054  
3237 TTGCTGACAGAGAAATCCGAACACTGGCTACCAAGAAATTCACGAATTCGCCGTTAAGTTG 3296  
3055 TTGCTGACAGAGAAATCCGAACACTGGCTACCAAGAAATTCACGAATTCGCCGTTAAGTTG 3114  
3297 CAGGATTTGACAGGTTCTGGAACACATGCTAATAAATTTGCTCABAATGCTTCTGCTGAT 3356  
3115 CAGGATTTGACAGGTTCTGGAACACATGCTAATAAATTTGCTCABAATGCTTCTGCTGAT 3174  
3357 ATCAGCAGCTAAATAATTTCCACCAACTCAGGAATCTCTATGATCCCAACCTGCCA 3416  
3175 ATCAGCAGCTAAATAATTTCCACCAACTCAGGAATCTCTATGATCCCAACCTGCCA 3234  
3417 CCGTCACTAAAGTCTAGTAACAACTGTAACCAAGTAACTGACAAATTCACAAAGCA 3476  
3235 CCGTCACTAAAGTCTAGTAACAACTGTAACCAAGTAACTGACAAATTCACAAAGCA 3294  
3477 TATAAGGACAAAAACAATAATAGTTTGAATCATGGGAGAGAAAGAAATCGCTTTTAAA 3536  
3295 TATAAGGACAAAAACAATAATAGTTTGAATCATGGGAGAGAAAGAAATCGCTTTTAAA 3354  
3537 ATGATTTCTGACCAACGTTTCTCATGTGGTGGCCAGTTGGATGACATAGAAAAACCCCT 3596  
3355 ATGATTTCTGACCAACGTTTCTCATGTGGTGGCCAGTTGGATGACATAGAAAAACCCCT 3414  
3597 AGGAAGTTTGTTCCTGTAATGACAACTGACCAATTCATAAAGATGCTCAGACAGTG 3656  
3415 AGGAAGTTTGTTCCTGTAATGACAACTGACCAATTCATAAAGATGCTCAGACAGTG 3474  
3657 AAGCTGTTCTCAGGAGCTTCTATGAAATCCATGTTCCCATACCTTCCCAATTTGAACCTG 3716  
3475 AAGCTGTTCTCAGGAGCTTCTATGAAATCCATGTTCCCATACCTTCCCAATTTGAACCTG 3534  
3717 CCAAGAGATATCAAAACCGTTTCTCATGATGATGATGATGATGATGATGATGATGATGAT 3776  
3535 CCAAGAGATATCAAAACCGTTTCTCATGATGATGATGATGATGATGATGATGATGATGAT 3594  
3777 CGAGACAAATGAAAGTTTGGAC 3799  
3595 CGAGACAAATGAAAGTAGTAGTC 3617

Patent No. 6642038  
GENERAL INFORMATION:  
APPLICANT: CANFIELD, WILLIAM M  
TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY  
FILE REFERENCE: 210119US00CNT  
CURRENT APPLICATION NUMBER: US/09/636,060C  
CURRENT FILING DATE: 2000-08-10  
PRIOR APPLICATION NUMBER: 60/153,831  
PRIOR FILING DATE: 1999-09-14  
NUMBER OF SEQ ID NOS: 52  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 21  
LENGTH: 3621  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-636-060C-21

Query Match 62.3%; Score 3485.4; DB 4; Length 3621;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 3492; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 297 AGCCGAGATCAATACATATGTTTGTGATTCCTATAGAGACAAATTTGCTGGAAAGTCC 356  
DB 115 AGCCGAGATCAATACATATGTTTGTGATTCCTATAGAGACAAATTTGCTGGAAAGTCC 174  
QY 357 TTTTCAGATCGGCTTTGCTGCCCATGCCGATGACGTTGTTTACACCTGGGTGAATGCG 416  
DB 175 TTTTCAGATCGGCTTTGCTGCCCATGCCGATGACGTTGTTTACACCTGGGTGAATGCG 234  
QY 417 ACAGATCTTTGAACTACTGAAGGAATCTACAGAGGTACAGAGCAACAGATCGAGAGAGCAG 476  
DB 235 ACAGATCTTTGAACTACTGAAGGAATCTACAGAGGTACAGAGCAACAGATCGAGAGAGCAG 294  
QY 477 AAAGCAATGAGAGAAATCTTGGGAAACACACAGGAACTTACTAAGAGAGGTGAGAG 536  
DB 295 AAAGCAATGAGAGAAATCTTGGGAAACACACAGGAACTTACTAAGAGAGGTGAGAG 354  
QY 537 CAGTTAGAGTGTTCCTTAACACACTGCAATTAAGGTGCCAATGCTGTACTGGACCCAGCC 596  
DB 355 CAGTTAGAGTGTTCCTTAACACACTGCAATTAAGGTGCCAATGCTGTACTGGACCCAGCC 414  
QY 597 CTGCCAGCAACATCACCTGAGAGAGTGCATCTCTTTATCTCTTCTTTTCTTTCTGCC 656  
DB 415 CTGCCAGCAACATCACCTGAGAGAGTGCATCTCTTTATCTCTTCTTTTCTTTCTGCC 474  
QY 657 AGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCAGTTGTTGTT 716  
DB 475 AGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCAGTTGTTGTT 534  
QY 717 TTTGACAGTACTAAGATGTTGAAGATGCCCACTGCACTGCTTAAAGAAATAGCAGA 776  
DB 535 TTTGACAGTACTAAGATGTTGAAGATGCCCACTGCACTGCTTAAAGAAATAGCAGA 594  
QY 777 CAGACAGTATGAGGGGTACTTGACACACATTAAGAGTCCCTGGATAGTCTAATG 836  
DB 595 CAGACAGTATGAGGGGTACTTGACACACATTAAGAGTCCCTGGATAGTCTAATG 654  
QY 837 CAAAGATTTGGCTTCTGAGTGGATTTCCACCAACATTTCAAGGAAACAAATCAACTAAAA 896  
DB 655 CAAAGATTTGGCTTCTGAGTGGATTTCCACCAACATTTCAAGGAAACAAATCAACTAAAA 714  
QY 897 ACATAATTTGCCAGAAATCTTCTCTTAAGTCAAACTGTTGCGAGTGTATTCAGAGCC 956  
DB 715 ACATAATTTGCCAGAAATCTTCTCTTAAGTCAAACTGTTGCGAGTGTATTCAGAGCC 774  
QY 957 AGTGTAGCGCTTCTAAAACTGAATAAACCCTGCAAGATTTTCAAGAAATGAATAGCAAACT 1016  
DB 775 AGTGTAGCGCTTCTAAAACTGAATAAACCCTGCAAGATTTTCAAGAAATGAATAGCAAACT 834  
QY 1017 AAGAAGACATGACCATGATGAGAAAGAACTGACCAATAGTCTGATATTTATG 1076  
DB 835 AAGAAGACATGACCATGATGAGAAAGAACTGACCAATAGTCTGATATTTATG 894









Db	1315	TCCTGGATTAGGATGGCTATTGTGCAAGGCTTGTAAATTCAGCGCTGGATTGGAT	1374
Qy	1557	GGTGGGGAATTGCTCTGGAACACAGTGGAGGAGTGCCTATATTGAGAGGTTGGAGTACT	1616
Db	1375	GGTGGGGAATTGCTCTGGAACACAGTGGAGGAGTGCCTATATTGAGAGGTTGGAGTACT	1434
Qy	1617	GGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGGTGAGGAAATAAACAAGTGTCTTTAC	1676
Db	1435	GGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGGTGAGGAAATAAACAAGTGTCTTTAC	1494
Qy	1677	TGTAATCAGGGAATGGCGAATTCCTGGCTCGCTGATTAAGTTCTGTGACCAAGCATGCAT	1736
Db	1495	TGTAATCAGGGAATGGCGAATTCCTGGCTCGCTGATTAAGTTCTGTGACCAAGCATGCAT	1554
Qy	1737	GTCTTGTCTGTGGGTTTGTAGCTGCGCACTGTGGGCAAGATCATTTTCATGAATTGTAT	1796
Db	1555	GTCTTGTCTGTGGGTTTGTAGCTGCGCACTGTGGGCAAGATCATTTTCATGAATTGTAT	1614
Qy	1797	AAAGTGATCTTCTCCCAACAGACTCACTATATTATTCCAAAGAGTGAATGCTGCT	1856
Db	1615	AAAGTGATCTTCTCCCAACAGACTCACTATATTATTCCAAAGAGTGAATGCTGCT	1674
Qy	1857	TATTTTCAGCTTTGCAAGTAGCCCAAGAGGTTGAAGTGGCTATAGTGACAATCCA	1916
Db	1675	TATTTTCAGCTTTGCAAGTAGCCCAAGAGGTTGAAGTGGCTATAGTGACAATCCA	1734
Qy	1917	ATAATTTCGACATGCTTCTTATTTCCCAACAGTGGGAAACCATCCACTCAATGACACAGT	1976
Db	1735	ATAATTTCGACATGCTTCTTATTTCCCAACAGTGGGAAACCATCCACTCAATGACACAGT	1794
Qy	1977	GGAAATGAATGCGCACCAATATACATTTAATCTCAGTTTCAAAATACAAACGATGAAGAG	2036
Db	1795	GGAAATGAATGCGCACCAATATACATTTAATCTCAGTTTCAAAATACAAACGATGAAGAG	1854
Qy	2037	TTCAAAATCAGATAACAGTGGAGTGGACAAAGGAGGACCAAAATCTGAATTTCTACG	2096
Db	1855	TTCAAAATCAGATAACAGTGGAGTGGACAAAGGAGGACCAAAATCTGAATTTCTACG	1914
Qy	2097	GCCAGAGGGTTACGAAATTTAGTTAGTCCCATACACTTTCTTCAGAGCGGAAATC	2156
Db	1915	GCCAGAGGGTTACGAAATTTAGTTAGTCCCATACACTTTCTTCAGAGCGGAAATC	1974
Qy	2157	CTTTTTCAGGATATCCCAAGAAAGACGTTTCCCGAGTTTAAAGACATGATCTTAAC	2216
Db	1975	CTTTTTCAGGATATCCCAAGAAAGACGTTTCCCGAGTTTAAAGACATGATCTTAAC	2034
Qy	2217	TCAACAGGAGGAGCCAGGAAGGTGAAATTTCCCTGGTAAATTTTCACTCTCTTCCA	2276
Db	2035	TCAACAGGAGGAGCCAGGAAGGTGAAATTTCCCTGGTAAATTTTCACTCTCTTCCA	2094
Qy	2277	AAAGACGCCAGTTGAGTCTCAATACCTTGGATTGCAACTGGAAATGGAACATCACT	2336
Db	2095	AAAGACGCCAGTTGAGTCTCAATACCTTGGATTGCAACTGGAAATGGAACATCACT	2154
Qy	2337	TTGAAAGGATACAAATTTGTCAGGTCAGCTTGTGAGATCAATTTCTGATGAATCAACAG	2396
Db	2155	TTGAAAGGATACAAATTTGTCAGGTCAGCTTGTGAGATCAATTTCTGATGAATCAACAG	2214
Qy	2397	CATGCTAAATAAATAAATCAAGCTATAATAACAGATGAACAAATGACAGTTTGGTGGCT	2456
Db	2215	CATGCTAAATAAATAAATCAAGCTATAATAACAGATGAACAAATGACAGTTTGGTGGCT	2274
Qy	2457	CCACAGGAAACACAGCTTCAATAAGCATCTTCCCAACAGCTTGAAGTGTCTGAAGA	2516
Db	2275	CCACAGGAAACACAGCTTCAATAAGCATCTTCCCAACAGCTTGAAGTGTCTGAAGA	2334
Qy	2517	TTGACAGGTTGATCTTTCTGCAAGTGAAGTGAATGAATGGTCAATGACAGGAGTACAG	2576
Db	2335	TTGACAGGTTGATCTTTCTGCAAGTGAAGTGAATGAATGGTCAATGACAGGAGTACAG	2394
Qy	2577	AATCCACCCCTGGACTTGGAGACACAGCAAGATTTAGAGTGAAGAACTCACACCCAA	2636

Db	2395	AATCCACCCCTGGACTTGGAGACCACAGCAAGATTTAGAGTGGAAACTCACACCCCAAAA	2454
Qy	2637	ACCATAGGCGGAAATGTGAAAAAGAAAGCCCCCATCTCTGATTGTCTTCTACCTGGAAAGC	2696
Db	2455	ACCATAGGCGGAAATGTGAAAAAGAAAGCCCCCATCTCTGATTGTCTTCTACCTGGAAAGC	2514
Qy	2697	CAGATGACAAAAAGAAAGAAATCACAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA	2756
Db	2515	CAGATGACAAAAAGAAAGAAATCACAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA	2574
Qy	2757	AATGCTGAAATCACATAGGCGTTACTTGAAGTGTACTTGGAGAAAGCTGCAGCATTTAC	2816
Db	2575	AATGCTGAAATCACATAGGCGTTACTTGAAGTGTACTTGGAGAAAGCTGCAGCATTTAC	2634
Qy	2817	ACAGATAGTTACTTGGGCTTTTGGCCATGGGAGAAAAAGAAAGAAAGAAAGAAAGAAAG	2876
Db	2635	ACAGATAGTTACTTGGGCTTTTGGCCATGGGAGAAAAAGAAAGAAAGAAAGAAAGAAAG	2694
Qy	2877	GAGGAGAGAGTCATTGAGACACAATTTGGCATACTTCTCAGTATGCAAAAAATCTCTGG	2936
Db	2695	GAGGAGAGAGTCATTGAGACACAATTTGGCATACTTCTCAGTATGCAAAAAATCTCTGG	2754
Qy	2937	AGGCAACTAAAGATACATTTGAGATTCCTCAGATATGTAAATTAATTTCTAAATAGC	2996
Db	2755	AGGCAACTAAAGATACATTTGAGATTCCTCAGATATGTAAATTAATTTCTAAATAGC	2814
Qy	2997	AAGTTTGGATTACATCGGGGAAAGTCCCTGCTCAGATGCTCAGATGATGACCCGAT	3056
Db	2815	AAGTTTGGATTACATCGGGGAAAGTCCCTGCTCAGATGCTCAGATGATGACCCGAT	2874
Qy	3057	GTATGCAAGAACTGCAAGATATGTTCCCTGAAAGAAATTTGACAGAGCGTCATTTCAAAA	3116
Db	2875	GTATGCAAGAACTGCAAGATATGTTCCCTGAAAGAAATTTGACAGAGCGTCATTTCAAAA	2934
Qy	3117	GTGCGCATTTCTGAGGATGAGATTTGCGCTTCTTTATTTTATTTATCTCATGAGTGCA	3176
Db	2935	GTGCGCATTTCTGAGGATGAGATTTGCGCTTCTTTATTTTATTTATCTCATGAGTGCA	2994
Qy	3177	GTGAGAGCCTGATATATCTCAAGTCTTTGATGAGTTGATACAGATCAATCTGCTGTC	3236
Db	2995	GTGAGAGCCTGATATATCTCAAGTCTTTGATGAGTTGATACAGATCAATCTGCTGTC	3054
Qy	3237	TTGCTGACAGAGAAATCCGAAACACTGCTGCTACAGAAATTCAGAACTGCGGTTAAGTTT	3296
Db	3055	TTGCTGACAGAGAAATCCGAAACACTGCTGCTACAGAAATTCAGAACTGCGGTTAAGTTT	3114
Qy	3297	CAGGATTTGACAGTCTGGAACACATGCTAAATAAATTTGCTCAAAATGCTTCTGCTGAT	3356
Db	3115	CAGGATTTGACAGTCTGGAACACATGCTAAATAAATTTGCTCAAAATGCTTCTGCTGAT	3174
Qy	3357	ATCAGCAGGCTTAAATAATTTCCACCAACTCAGGAATCTTACTATGATCCCAACCTGCCA	3416
Db	3175	ATCAGCAGGCTTAAATAATTTCCACCAACTCAGGAATCTTACTATGATCCCAACCTGCCA	3234
Qy	3417	CCGCTCACTAAAGTCTAGTAAACAACTGTAAACCCAGTAACTGACAAAATCCCAAGCA	3476
Db	3235	CCGCTCACTAAAGTCTAGTAAACAACTGTAAACCCAGTAACTGACAAAATCCCAAGCA	3294
Qy	3477	TATTAAGACAAAAAATAATAGGTTTGAATCATGGAAGAAAGAAATGCGTTTTTAA	3536
Db	3295	TATTAAGACAAAAAATAATAGGTTTGAATCATGGAAGAAAGAAATGCGTTTTTAA	3354
Qy	3537	ATGATTCGTACCAACGTTTCTCATGCTGGTGGCCAGTTGGATGACATGAAGAAAAACCT	3596
Db	3355	ATGATTCGTACCAACGTTTCTCATGCTGGTGGCCAGTTGGATGACATGAAGAAAAACCT	3414
Qy	3597	AGGAGTTTGTGCTTGCCTGAATGACAACTTGACCACAATATAAAGATGCTCAGACAGTG	3656
Db	3415	AGGAGTTTGTGCTTGCCTGAATGACAACTTGACCACAATATAAAGATGCTCAGACAGTG	3474
Qy	3657	AAGGCTTCTCAGGAGCTTCTATGAATFCAATGTTCCCACTACCTTCCCAATTTGAACCTG	3716
Db	3475	AAGGCTTCTCAGGAGCTTCTATGAATFCAATGTTCCCACTACCTTCCCAATTTGAACCTG	3534



QY 1977 GGAATGAATGCCACACAAATACATTTTAACTCTCAGCTTTCAAATAAAGCATGAAGAG 2036  
DB 1795 GGAATGAATGCCACACAAATACATTTTAACTCTCAGCTTTCAAATAAAGCATGAAGAG 1854  
QY 2037 TTCAAAATGCAGATTAACAGTGGAGTGACACAAAGGGAGGACCAAAATCTAATCTACG 2096  
DB 1855 TTCAAAATGCAGATTAACAGTGGAGTGACACAAAGGGAGGACCAAAATCTAATCTACG 1914  
QY 2097 GCCCAAGAGGGTTACGAAATTTAGTTAGTCCATTAACACATTTCTCCAGAGCGGAAATC 2156  
DB 1915 GCCCAAGAGGGTTACGAAATTTAGTTAGTCCATTAACACATTTCTCCAGAGCGGAAATC 1974  
QY 2157 CTTTTTGGAGTATTTCCCAAGAAAGAAAGCTTTCCCGAAGTTTAAAGAGACATGATTTAA 2216  
DB 1975 CTTTTTGGAGTATTTCCCAAGAAAGAAAGCTTTCCCGAAGTTTAAAGAGACATGATTTAA 2034  
QY 2217 TCACAAGAGAGAGCCAGAGAGGTGAATAATTCCTCTGGTAAATATTTCTCACTCTTCCA 2276  
DB 2035 TCACAAGAGAGAGCCAGAGAGGTGAATAATTCCTCTGGTAAATATTTCTCACTCTTCCA 2094  
QY 2277 AAAGAGCCCGAGTTGAGTCTCAATACCTTTGGATTTGCAACTGGAACTGGAGACATCACT 2336  
DB 2095 AAAGAGCCCGAGTTGAGTCTCAATACCTTTGGATTTGCAACTGGAACTGGAGACATCACT 2154  
QY 2337 TTGAAAGGATACAAATTTGTCCAAAGTCAGCTTTGCTGAGATCAATTTCTGATCAATCTCAG 2396  
DB 2155 TTGAAAGGATACAAATTTGTCCAAAGTCAGCTTTGCTGAGATCAATTTCTGATCAATCTCAG 2214  
QY 2397 CATGCTAAATAAATAATCAAGCTATATAACAGATGAAGAAATAAGTGAAGTTGTGGCT 2456  
DB 2215 CATGCTAAATAAATAATCAAGCTATATAACAGATGAAGAAATAAGTGAAGTTGTGGCT 2274  
QY 2457 CCACAGGAAACACAGTTTCATAAAGAGCTTTGCCAAACAGCTTAGGAGTGTCTCAAGA 2516  
DB 2275 CCACAGGAAACACAGTTTCATAAAGAGCTTTGCCAAACAGCTTAGGAGTGTCTCAAGA 2334  
QY 2517 TTGCGAGAGTTGACTTTTCTCGAGTGTGTAAAGTGAATGGTTCATGACACAGGTCAG 2576  
DB 2335 TTGCGAGAGTTGACTTTTCTCGAGTGTGTAAAGTGAATGGTTCATGACACAGGTCAG 2394  
QY 2577 AATCCACCCCTGGACTTGGAGACCAACAGCAAGTTTAGAGTGAAGAACTCACACCAGAAA 2636  
DB 2395 AATCCACCCCTGGACTTGGAGACCAACAGCAAGTTTAGAGTGAAGAACTCACACCAGAAA 2454  
QY 2637 ACCATAGGGGAAATGTGACAAAGAAAGAGCCCATCTCTGATTTGCCACTGGAAAGC 2696  
DB 2455 ACCATAGGGGAAATGTGACAAAGAAAGAGCCCATCTCTGATTTGCCACTGGAAAGC 2514  
QY 2697 CAGATGACAAAGAAAGAAATTCACAGGAAAGAAAGAGAAACAGTAGAATGGAGGAA 2756  
DB 2515 CAGATGACAAAGAAAGAAATTCACAGGAAAGAAAGAGAAACAGTAGAATGGAGGAA 2574  
QY 2757 AATGCTGAAATTCACATAGCGTTACTGAAGTGTACTTTGGAAAGAAAGCTGACAGATTAC 2816  
DB 2575 AATGCTGAAATTCACATAGCGTTACTGAAGTGTACTTTGGAAAGAAAGCTGACAGATTAC 2634  
QY 2817 ACAGATAGTTACTTGGGCTTTTGGCCATGGGAGAAAGAAATTTTCCAGATCTTCTC 2876  
DB 2635 ACAGATAGTTACTTGGGCTTTTGGCCATGGGAGAAAGAAATTTTCCAGATCTTCTC 2694  
QY 2877 GACGAGAGAGTCAATGAGACACAAATTTGGCATCTTCACTGATAGCAAAATCTGGG 2936  
DB 2695 GACGAGAGAGTCAATGAGACACAAATTTGGCATCTTCACTGATAGCAAAATCTGGG 2754  
QY 2937 AGGCAACTAAAGATACATTTGAGATTTCCCTCAGATATGTAATAAATAATTTCTAAATAGC 2996  
DB 2755 AGGCAACTAAAGATACATTTGAGATTTCCCTCAGATATGTAATAAATAATTTCTAAATAGC 2814  
QY 2997 AAGTTTGGATTCACATCGGGAAAGTCCCTGCTCAGATGCTCAGATGATGACCGGATT 3056  
DB 2815 AAGTTTGGATTCACATCGGGAAAGTCCCTGCTCAGATGCTCAGATGATGACCGGATT 2874  
QY 3057 GTTATGCAAGAACTGCAAGATATGTTTCCCTGGAAGATTTTGAAGAGCTCATTTACAAA 3116

DB 2875 GTTATGCAAGAACTGCAAGATATGTTTCCCTGGAAGATTTGACAAGAGCTCATTTACAAA 2934  
QY 3117 GTGCGCAATTTCTGAGGATATGAGTTTGGCTTCTCTTATTTTATTATCTCATGAGTGCA 3176  
DB 2935 GTGCGCAATTTCTGAGGATATGAGTTTGGCTTCTCTTATTTTATTATCTCATGAGTGCA 2994  
QY 3177 GTGCGCACTGGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTGTC 3236  
DB 2995 GTGCGCACTGGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTGTC 3054  
QY 3237 TTGCTGACAGAGAAATTCGAAACACACACCTGGCTACAGAAATTCAGAACTGGCGTTAAGTTG 3296  
DB 3055 TTGCTGACAGAGAAATTCGAAACACACCTGGCTACAGAAATTCAGAACTGGCGTTAAGTTG 3114  
QY 3297 CAGGNTTGCAGAGTCTGGAAACACATCTTAATAATTTGCTCAAAATGCTTCTCGCTGAT 3356  
DB 3115 CAGGNTTGCAGAGTCTGGAAACACATCTTAATAATTTGCTCAAAATGCTTCTCGCTGAT 3174  
QY 3357 ATCAGCAGGCTAAATAATATTTCCACCAACTCAGGAATCTCTATCTATGATCCCAACCTGCCA 3416  
DB 3175 ATCAGCAGGCTAAATAATATTTCCACCAACTCAGGAATCTCTATCTATGATCCCAACCTGCCA 3234  
QY 3417 CCGGTCTCTAAAAGTCTAGTAAACAACTGTAAACCCAGTAACTGACAAAATTCACAAAGCA 3476  
DB 3235 CCGGTCTCTAAAAGTCTAGTAAACAACTGTAAACCCAGTAACTGACAAAATTCACAAAGCA 3294  
QY 3477 TATAAGGACAAACAAATATAGTGTGTTGAGTGTGAGTGTGATGACATAAGAAAAACCT 3536  
DB 3295 TATAAGGACAAACAAATATAGTGTGTTGAGTGTGAGTGTGATGACATAAGAAAAACCT 3354  
QY 3537 ATGATTCGTACCAACCGTTTCTCATGTGTTGGCCAGTTGGATGACATAAGAAAAACCT 3596  
DB 3355 ATGATTCGTACCAACCGTTTCTCATGTGTTGGCCAGTTGGATGACATAAGAAAAACCT 3414  
QY 3597 AGGAGTTTGTTCCTGCTGAATGACCAATGACCAATCATAAAGATGCTCAGACAGTG 3656  
DB 3415 AGGAGTTTGTTCCTGCTGAATGACCAATGACCAATCATAAAGATGCTCAGACAGTG 3474  
QY 3657 AAGGCTGTCTCAGGGACTTCTATGAATCCATGTTCCTCATACCTTCCCAATTTGAACTG 3716  
DB 3475 AAGGCTGTCTCAGGGACTTCTATGAATCCATGTTCCTCATACCTTCCCAATTTGAACTG 3534  
QY 3717 CCAAGAGAGTATCGAAACCGTTTCTCATATGATGAGCTGAGGATGAGGGCTTAT 3776  
DB 3535 CCAAGAGAGTATCGAAACCGTTTCTCATATGATGAGCTGAGGATGAGGGCTTAT 3594  
QY 3777 CGAGACAAATTTGAAGTTTGGAC 3799  
DB 3595 CGAGACAAATTTGAAGTTAGTAGTC 3617

Search completed: November 22, 2004, 10:44:27  
Job time : 434 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 22, 2004, 06:14:52 ; Search time 2612 Seconds  
(without alignments)  
11578.897 Million cell updates/sec

Title: US-10-023-888-3

Perfect score: 5597

Sequence: 1 cggagcgcagcggcgctccg.....aaaagttaattttgaaa 5597

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3627888 seqs, 2701811610 residues

Total number of hits satisfying chosen parameters: 7255776

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 19: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*
- 20: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 21: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	5597	100.0	5597	9	US-09-895-072-4
2	5597	100.0	5597	9	US-09-986-552-4
3	5597	100.0	5597	15	US-10-023-888-3
4	5597	100.0	5597	15	US-10-023-889-3
5	5597	100.0	5597	15	US-10-023-890-3
6	5597	100.0	5597	15	US-10-024-197-3
7	5597	100.0	5597	15	US-10-023-894-3
8	5597	100.0	5597	15	US-10-306-686-4
9	3761	67.2	3783	9	US-09-895-072-20
10	3761	67.2	3783	9	US-09-986-552-20
11	3761	67.2	3783	15	US-10-306-686-20
12	3485.4	62.3	3621	9	US-09-895-072-21
					Sequence 4, Appli
					Sequence 4, Appli
					Sequence 3, Appli
					Sequence 3, Appli
					Sequence 3, Appli
					Sequence 3, Appli
					Sequence 4, Appli
					Sequence 20, Appli
					Sequence 20, Appli
					Sequence 21, Appli

13	3485.4	62.3	3621	9	US-09-986-552-21	Sequence 21, Appli
14	3485.4	62.3	3621	15	US-10-306-686-21	Sequence 21, Appli
15	3465.6	61.9	3600	15	US-10-023-888-1	Sequence 1, Appli
16	3465.6	61.9	3600	15	US-10-023-889-1	Sequence 1, Appli
17	3465.6	61.9	3600	15	US-10-023-890-1	Sequence 1, Appli
18	3465.6	61.9	3600	15	US-10-024-197-1	Sequence 1, Appli
19	3465.6	61.9	3600	15	US-10-023-894-1	Sequence 1, Appli
20	2820.8	50.4	5229	9	US-09-895-072-16	Sequence 16, Appli
21	2820.8	50.4	5229	9	US-09-986-552-16	Sequence 16, Appli
22	2820.8	50.4	5229	15	US-10-023-888-8	Sequence 8, Appli
23	2820.8	50.4	5229	15	US-10-023-889-8	Sequence 8, Appli
24	2820.8	50.4	5229	15	US-10-023-890-8	Sequence 8, Appli
25	2820.8	50.4	5229	15	US-10-024-197-8	Sequence 8, Appli
26	2820.8	50.4	5229	15	US-10-023-894-8	Sequence 8, Appli
27	2820.8	50.4	5229	15	US-10-306-686-16	Sequence 16, Appli
28	2613	46.7	8811	15	US-10-120-801-25	Sequence 25, Appli
29	2597.4	46.2	2603	15	US-10-094-749-753	Sequence 753, Appli
30	1977.2	35.3	6447	15	US-10-172-118-53	Sequence 59, Appli
31	1977.2	35.3	6447	16	US-10-342-887-59	Sequence 59, Appli
32	1105.2	19.7	1194	15	US-10-029-386-22697	Sequence 22697, A
33	1104.8	19.7	1192	9	US-09-864-761-10195	Sequence 10195, A
34	1103.4	19.7	1125	9	US-09-864-761-26835	Sequence 26835, A
35	724.8	12.9	728	15	US-10-139-794-170	Sequence 170, Appli
36	415.4	7.4	455	10	US-09-918-995-25442	Sequence 25442, A
37	377.2	6.7	459	10	US-09-918-995-27675	Sequence 27675, A
38	339	6.1	572	11	US-09-969-034-612	Sequence 612, Appli
39	301.4	5.4	572	15	US-10-029-386-8979	Sequence 8979, A
40	256.4	4.6	259	16	US-10-305-720-402	Sequence 402, Appli
41	256	4.6	337	13	US-10-079-623-97	Sequence 97, Appli
42	255	4.6	601	9	US-09-833-381-2047	Sequence 2047, Appli
43	238.2	4.3	44848	9	US-09-776-874A-42	Sequence 42, Appli
44	238.2	4.3	44848	9	US-09-988-113-42	Sequence 42, Appli
45	238.2	4.3	44848	15	US-10-341-582-42	Sequence 42, Appli

ALIGNMENTS

RESULT 1

US-09-895-072-4  
; Sequence 4, Application US/09895072  
; Patent No. US2002002550A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM M  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 21019USCCONT  
; CURRENT APPLICATION NUMBER: US/09/895,072  
; CURRENT FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; PRIOR APPLICATION NUMBER: US 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 5597  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-895-072-4

Query Match	100.0%;	Score 5597;	DB 9;	Length 5597;
Best local Similarity	100.0%;	Pred. No. 0;		
Matches 5597;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
Qy	1	CGGAGCGGAGCGGCGCTCGTCCCGAGAGTGCATGAGCGGCGCCCGGAGGCTGTGACC	60	
Db	1	CGGAGCGGAGCGGCGCTCGTCCCGAGAGTGCATGAGCGGCGCCCGGAGGCTGTGACC	60	
Qy	61	TGGCGCGGCGGCGCCCGAGCGGCGCCCGTGAATGCGCGGCTCGCTGAGCGCGCGCGCGCGC	120	
Db	61	TGGCGCGGCGGCGCCCGAGCGGCGCCCGTGAATGCGCGCTCGCTGAGCGCGCGCGCGCGC	120	
Qy	121	GGCGGCTCAGGCTCTCGCGGCGCGTGGCGGTGAAGGGGTGATGCTTCAAGCTCC	180	

121 GCGCGCTCAGGCTCCTCGGGCGGTGGCGTGGCGTGAAGGGGTGATGCTGTCAAGCTCC 180  
181 TGCAGAGACAAACCTATACCTGCTGTCGCCACAGTATGGCTCTACGTGTCTTTCTGG 240  
181 TGCAGAGACAAACCTATACCTGCTGTCGCCACAGTATGGCTCTACGTGTCTTTCTGG 240  
241 GCGTGTGTGTCACCATGCTCTCGGCTTCCAGTTCGAGAGGTGGTTCGGAATGGAGCC 300  
241 GCGTGTGTGTCACCATGCTCTCGGCTTCCAGTTCGAGAGGTGGTTCGGAATGGAGCC 300  
301 GAGATCAATACCATGTTTGTGTTGATTCCTATAGAGCAATATTTGCTGGAAGTCTTTTC 360  
301 GAGATCAATACCATGTTTGTGTTGATTCCTATAGAGCAATATTTGCTGGAAGTCTTTTC 360  
361 AGAATCGGCTTTGTCTGCCCATGCCGATTGACGTTGTTTACACCTGGGTGAATGGCACAG 420  
361 AGAATCGGCTTTGTCTGCCCATGCCGATTGACGTTGTTTACACCTGGGTGAATGGCACAG 420  
421 ATCTTGAACTACTGAAGAACTACAGCAGGTGACAGAACAGATGGAGGGAGCGAAG 480  
421 ATCTTGAACTACTGAAGAACTACAGCAGGTGACAGAACAGATGGAGGGAGCGAAG 480  
481 CAATGAGAGAAATCCTTGGGAAAAACAACGGAACCTACTAAGAAGAGTGAGAAGCAGT 540  
481 CAATGAGAGAAATCCTTGGGAAAAACAACGGAACCTACTAAGAAGAGTGAGAAGCAGT 540  
541 TAGAGTGTGCTTAACACACTGCAATTAAGTGGCAATGCTTGTACTGGACCCAGCCCTGC 600  
541 TAGAGTGTGCTTAACACACTGCAATTAAGTGGCAATGCTTGTACTGGACCCAGCCCTGC 600  
601 CAGCAACATCACCCTGAAGACGTCGATCTCTTTTATCTCTTTTCTTTTCTGCGCAGTG 660  
601 CAGCAACATCACCCTGAAGACGTCGATCTCTTTTATCTCTTTTCTTTTCTGCGCAGTG 660  
661 ACATTTTCAATGTTGCAAAACCAAAAAACCTTCTACCAATGCTCAGTTGTTGTTTGT 720  
661 ACATTTTCAATGTTGCAAAACCAAAAAACCTTCTACCAATGCTCAGTTGTTGTTTGT 720  
721 ACAGTACTAAGAGTGTGAAGATGCCACCTGACCTGACCTTAAAGGAATAGCAGACAGA 780  
721 ACAGTACTAAGAGTGTGAAGATGCCACCTGACCTGACCTTAAAGGAATAGCAGACAGA 780  
781 CAGTATGAGGGGGTACTTGACACACAGATAAAGAGTCCCTGGATTAGTCTAATGCAAG 840  
781 CAGTATGAGGGGGTACTTGACACACAGATAAAGAGTCCCTGGATTAGTCTAATGCAAG 840  
841 ATTTGGCTTCTCAGTGGATTCCACACATTCAGGAAACAAATCAACTAAACAA 900  
841 ATTTGGCTTCTCAGTGGATTCCACACATTCAGGAAACAAATCAACTAAACAA 900  
901 AATTGCCAGAAAATCTTTCTCTAAAGTCAAACTGTTGCAAGTGTATTTCAGAGCCAGTG 960  
901 AATTGCCAGAAAATCTTTCTCTAAAGTCAAACTGTTGCAAGTGTATTTCAGAGCCAGTG 960  
961 TAGCGCTTCTAAAACCTGAATAACCCCAAGGATTTTCAAGAAATGAATAAGCAAACTAAG 1020  
961 TAGCGCTTCTAAAACCTGAATAACCCCAAGGATTTTCAAGAAATGAATAAGCAAACTAAG 1020  
1021 AGAATGACCATTTGATGGAAGAACTGACCAATAGTCTGATATTTATTTATGGGATC 1080  
1021 AGAATGACCATTTGATGGAAGAACTGACCAATAGTCTGATATTTATTTATGGGATC 1080  
1081 TGAGCGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCGAGTCTGTTTCAAGATA 1140  
1081 TGAGCGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCGAGTCTGTTTCAAGATA 1140  
1141 ACGAAGAACTGAGGTACTTATGCGATCTATCGAGGCGATGACCATGGGTTTCSAATA 1200  
1141 ACGAAGAACTGAGGTACTTATGCGATCTATCGAGGCGATGACCATGGGTTTCSAATA 1200  
1201 TTTTCAATGTCACCAACGGGAGATTCATCTCTGCTGAACTTGAACATTCCTCGAGTGA 1260

1201 TTTTCAATGTCACCAACGGGAGATTCATCTCTGCTGAACTTGAACAATCCTCGAGTGA 1260  
1261 CAATAGTAACACACACAGAGATGTTTTTCGAAATTTGAGCCACTTGTCTACCTTTAGTTTCA 1320  
1261 CAATAGTAACACACACAGAGATGTTTTTCGAAATTTGAGCCACTTGTCTACCTTTAGTTTCA 1320  
1321 CTGCTATTGAAAGTGCATTCATCGCATCGAAGGGCTGTCCGAGAGTATTTTACCTAA 1380  
1321 CTGCTATTGAAAGTGCATTCATCGCATCGAAGGGCTGTCCGAGAGTATTTTACCTAA 1380  
1381 ATGATGATGCTCATGTTTGGAGAGATGTCGCGCAGATGATTTTACAGTCACTCCAAAG 1440  
1381 ATGATGATGCTCATGTTTGGAGAGATGTCGCGCAGATGATTTTACAGTCACTCCAAAG 1440  
1441 GCCAGAGGTTTATTTGACATGGCTGTGCCAAAACCTGTGCCGAGGGTGCACAGTTCT 1500  
1441 GCCAGAGGTTTATTTGACATGGCTGTGCCAAAACCTGTGCCGAGGGTGCACAGTTCT 1500  
1501 GGATTAAGAGTGGCTATTGTCACAGGCTTGTAAATAATTCAGCTGCGATTTGGAGTGGT 1560  
1501 GGATTAAGAGTGGCTATTGTCACAGGCTTGTAAATAATTCAGCTGCGATTTGGAGTGGT 1560  
1561 GGGATTGCTCTGGAACACAGTGGAGGAGTGCCTATTTTGCAGGAGGTGGAGTACTGGGA 1620  
1561 GGGATTGCTCTGGAACACAGTGGAGGAGTGCCTATTTTGCAGGAGGTGGAGTACTGGGA 1620  
1621 GTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAACAGTGTCTCTTACTGTA 1680  
1621 GTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAACAGTGTCTCTTACTGTA 1680  
1681 ATCAGGAGTGTGCGAATTTCTGGCTCGCTGATGATAGTTCTGTGACCAAGCATGCAATGCT 1740  
1681 ATCAGGAGTGTGCGAATTTCTGGCTCGCTGATGATAGTTCTGTGACCAAGCATGCAATGCT 1740  
1741 TGTCTGTGGGTTTGAATGCTGCGCATGCTGGGCAAGATCATTTTTCATGAAATGTATAAG 1800  
1741 TGTCTGTGGGTTTGAATGCTGCGCATGCTGGGCAAGATCATTTTTCATGAAATGTATAAG 1800  
1801 TGATCTTCTCCCAACCCAGACTCACTATATTTATTTCCAAAAGTGAATGCTGCTTATT 1860  
1801 TGATCTTCTCCCAACCCAGACTCACTATATTTATTTCCAAAAGTGAATGCTGCTTATT 1860  
1861 TCAGCTTTGCAAGAGTAGCAGGAGTGAAGTGGCTTATAGTGAACAATCCAATAA 1920  
1861 TCAGCTTTGCAAGAGTAGCAGGAGTGAAGTGGCTTATAGTGAACAATCCAATAA 1920  
1921 TTTGCAATGCTTATTTGCGCAAGTGAAGAAACCATCCCTCATATGACAGTGGAA 1980  
1921 TTTGCAATGCTTATTTGCGCAAGTGAAGAAACCATCCCTCATATGACAGTGGAA 1980  
1981 TGAATGCCACCAACATATATTTTATCTCAGCTTTCAAAATACAAACGATGAAGAGTTCA 2040  
1981 TGAATGCCACCAACATATATTTTATCTCAGCTTTCAAAATACAAACGATGAAGAGTTCA 2040  
2041 AAATCAGATTAACAGTGAAGTGAACACAGGGAGGACCAAACTGAAATTTCTAGGCCC 2100  
2041 AAATCAGATTAACAGTGAAGTGAACACAGGGAGGACCAAACTGAAATTTCTAGGCCC 2100  
2101 AGAAGGGTTAGCAAAATTTAGTTAGTCCCATAAACACTTTCTTCCAGAGGCGGAAATCTTT 2160  
2101 AGAAGGGTTAGCAAAATTTAGTTAGTCCCATAAACACTTTCTTCCAGAGGCGGAAATCTTT 2160  
2161 TTGAGGATATTTCCAAAGAAAACGCTTTCCGAAAGTTTAAAGACATGATGTTTAACTCAA 2220  
2161 TTGAGGATATTTCCAAAGAAAACGCTTTCCGAAAGTTTAAAGACATGATGTTTAACTCAA 2220  
2221 CAAAGAGAGCCAGGAGAGGTGAATAATTTCCCTTGGTAAATTTCTCTCTTCCAAAAG 2280  
2221 CAAAGAGAGCCAGGAGAGGTGAATAATTTCCCTTGGTAAATTTCTCTCTTCCAAAAG 2280  
2281 ACGCCAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAAATGGAACATGAGACATCTTTGA 2340  
2281 ACGCCAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAAATGGAACATGAGACATCTTTGA 2340



QY 2341 AAGGATACAAATTTGTCCAAAGTCAGCTTGTGAGATCAATTTCTGATGAACCTCACAGCATG 2400  
DB 2341 AAGGATACAAATTTGTCCAAAGTCAGCTTGTGAGATCAATTTCTGATGAACCTCACAGCATG 2400  
QY 2401 CTAAATTAATAAATCAAGCTATAATAACAGATGAACAAATGACAGATTTGGTGGCTCCAC 2460  
DB 2401 CTAAATTAATAAATCAAGCTATAATAACAGATGAACAAATGACAGATTTGGTGGCTCCAC 2460  
QY 2461 AGGAAAAACAGGTTTCATATAAGCATCTTGCACAAACAGCTTAGGAGTGTCTGAAGATTCG 2520  
DB 2461 AGGAAAAACAGGTTTCATATAAGCATCTTGCACAAACAGCTTAGGAGTGTCTGAAGATTCG 2520  
QY 2521 AGAGTTGATCTTTCTGCAAGTGTAAAGTGAATGGTCAATGACAGGCTCAGATC 2580  
DB 2521 AGAGTTGATCTTTCTGCAAGTGTAAAGTGAATGGTCAATGACAGGCTCAGATC 2580  
QY 2581 CACCCCTGGACTTGGAGACCAAGCAAGATTTAGAGTGGAACTCACACCCAAAAACCA 2640  
DB 2581 CACCCCTGGACTTGGAGACCAAGCAAGATTTAGAGTGGAACTCACACCCAAAAACCA 2640  
QY 2641 TAGGCGGAAATGTGACAAAAAGAAAGCCCATCTCTGATTTCCACTGGAAAGCCAGA 2700  
DB 2641 TAGGCGGAAATGTGACAAAAAGAAAGCCCATCTCTGATTTCCACTGGAAAGCCAGA 2700  
QY 2701 TGACAAAAGAAAGAAATCACAGGAAAGAAAGAGACAGTAGAATGGAGGAAATG 2760  
DB 2701 TGACAAAAGAAAGAAATCACAGGAAAGAAAGAGACAGTAGAATGGAGGAAATG 2760  
QY 2761 CTGAAATACATAGGCGTTACTGAAAGTGTACTTGGAAAGAAAGTGCAGCATTACACAG 2820  
DB 2761 CTGAAATACATAGGCGTTACTGAAAGTGTACTTGGAAAGAAAGTGCAGCATTACACAG 2820  
QY 2821 ATAGTTACTTGGGCTTTTGGCAAGGAGAAAGAAAGATTTTCCAGATCTTCTCGACG 2880  
DB 2821 ATAGTTACTTGGGCTTTTGGCAAGGAGAAAGAAAGATTTTCCAGATCTTCTCGACG 2880  
QY 2881 AAGAAGAGTCATTGAAGACACAATGGCATCTTCCACTGATAGCAAAATATCTGGGAGC 2940  
DB 2881 AAGAAGAGTCATTGAAGACACAATGGCATCTTCCACTGATAGCAAAATATCTGGGAGC 2940  
QY 2941 AACTAAAAAGATACATTTGCGATTCCTCAGATATGTAATAAATAAATTTCTAAATAGCAAGT 3000  
DB 2941 AACTAAAAAGATACATTTGCGATTCCTCAGATATGTAATAAATAAATTTCTAAATAGCAAGT 3000  
QY 3001 TTGGATTCACATCGCGGAAAGTCCCTGCTCAGATGCTCAGATGTTGACCGGATGTTA 3060  
DB 3001 TTGGATTCACATCGCGGAAAGTCCCTGCTCAGATGCTCAGATGTTGACCGGATGTTA 3060  
QY 3061 TGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAGACGTCATTTTCAAAAGTGC 3120  
DB 3061 TGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAGACGTCATTTTCAAAAGTGC 3120  
QY 3121 GCCATTTGAGGATATGAGTTTGCCTTCTCTATTATTTATATCTCAATGAGTGCAGTGC 3180  
DB 3121 GCCATTTGAGGATATGAGTTTGCCTTCTCTATTATTTATATCTCAATGAGTGCAGTGC 3180  
QY 3181 AGCCACTGAATATCTCAAGTCTTTGATGAGTTGATACAGATCAATCTGCTGCTTGT 3240  
DB 3181 AGCCACTGAATATCTCAAGTCTTTGATGAGTTGATACAGATCAATCTGCTGCTTGT 3240  
QY 3241 CTGACAGAGAAATCCGAACACTGGCTACCAAGATTTACGAACTGCGGTTAAGTTTGCAGG 3300  
DB 3241 CTGACAGAGAAATCCGAACACTGGCTACCAAGATTTACGAACTGCGGTTAAGTTTGCAGG 3300  
QY 3301 ATTTGACAGGTTGGAACACATGCTTAATAATGCTCAAAAATGCTTCTGCTGATATCA 3360  
DB 3301 ATTTGACAGGTTGGAACACATGCTTAATAATGCTTCAAAAATGCTTCTGCTGATATCA 3360  
QY 3361 CGCAGCTAAATATTTCCACCACTCAGGAATCTTACTATGATCCCACTGCGACCGG 3420  
DB 3361 CGCAGCTAAATATTTCCACCACTCAGGAATCTTACTATGATCCCACTGCGACCGG 3420

QY 3421 TCCTAAAAAGCTAGTAACAAACCTGTAACCAAGTAACCTGACAAAAATCCCAAGACATATA 3480  
DB 3421 TCCTAAAAAGCTAGTAACAAACCTGTAACCAAGTAACCTGACAAAAATCCCAAGACATATA 3480  
QY 3481 AGGACAAAAACAAATATAGGTTTGAATCATGGGAGAAAGAAATCGCTTTTAAATGA 3540  
DB 3481 AGGACAAAAACAAATATAGGTTTGAATCATGGGAGAAAGAAATCGCTTTTAAATGA 3540  
QY 3541 TTGCTACCAAGCTTTCTCATGTGTTGGCCAGTTGGATGACATAAAGAAAAACCCCTAGGA 3600  
DB 3541 TTGCTACCAAGCTTTCTCATGTGTTGGCCAGTTGGATGACATAAAGAAAAACCCCTAGGA 3600  
QY 3601 AGTTTGTTCCTCGAATGACAACTTGAACCAATCATAAAGATGCTCAGACAGTGAAGG 3660  
DB 3601 AGTTTGTTCCTCGAATGACAACTTGAACCAATCATAAAGATGCTCAGACAGTGAAGG 3660  
QY 3661 CTGTTCTCAGGACCTTCTATGAACTCCATGTTCCCATACCTTCCCAATTTGAACTGCCAA 3720  
DB 3661 CTGTTCTCAGGACCTTCTATGAACTCCATGTTCCCATACCTTCCCAATTTGAACTGCCAA 3720  
QY 3721 GAGAGTATCGAAACCCGTTTCTTCAATGATGAGCTGAGGAAATGGAGGCTTATCGAG 3780  
DB 3721 GAGAGTATCGAAACCCGTTTCTTCAATGATGAGCTGAGGAAATGGAGGCTTATCGAG 3780  
QY 3781 ACAATTTGAAGTTTGGACCCATTTGTACTAGCAACATTTGATTTTACTATATTCT 3840  
DB 3781 ACAATTTGAAGTTTGGACCCATTTGTACTAGCAACATTTGATTTTACTATATTCT 3840  
QY 3841 CATTTTGTGCTGAGCAGTTAAATGCACTTAAGCGGAAGATATTTCCAGAAAGGAGATAC 3900  
DB 3841 CATTTTGTGCTGAGCAGTTAAATGCACTTAAGCGGAAGATATTTCCAGAAAGGAGATAC 3900  
QY 3901 ACAAGAGCTAGTCCCAATCGAATCAGATATAGAGATCTTCATTTGAAACCACTCTA 3960  
DB 3901 ACAAGAGCTAGTCCCAATCGAATCAGATATAGAGATCTTCATTTGAAACCACTCTA 3960  
QY 3961 CCTCAGCATTTACTGAGCATTTTAAACCTCAGCTTCCAGAGATGCTTTGTGATGTGAT 4020  
DB 3961 CCTCAGCATTTACTGAGCATTTTAAACCTCAGCTTCCAGAGATGCTTTGTGATGTGAT 4020  
QY 4021 GCTTAGCAGTTTGGCCCGAAGAGAAATATCCAGTACCATGCTGTTTTGTGGCATGAA 4080  
DB 4021 GCTTAGCAGTTTGGCCCGAAGAGAAATATCCAGTACCATGCTGTTTTGTGGCATGAA 4080  
QY 4081 TATAGCCCATGCTAGTAAGTAATTTTAAACCAACCCACTGAAACCTTGTGTGCGACAGC 4140  
DB 4081 TATAGCCCATGCTAGTAAGTAATTTTAAACCAACCCACTGAAACCTTGTGTGCGACAGC 4140  
QY 4141 TCTGAACTGATTTTACTTTTAAAGAAATTTGCTCATGAGCCTGTCATCTTTTATAAAAA 4200  
DB 4141 TCTGAACTGATTTTACTTTTAAAGAAATTTGCTCATGAGCCTGTCATCTTTTATAAAAA 4200  
QY 4201 GGCTCACTGACAAAGACAGCTGTTAATTTCCCAAGCAATCATTTGCACTAACTTTAT 4260  
DB 4201 GGCTCACTGACAAAGACAGCTGTTAATTTCCCAAGCAATCATTTGCACTAACTTTAT 4260  
QY 4261 TAGGAGAGCCTATGCGAGCTGGAGTGTGTAAGAGGCTCCAGTCTTTGCAATTCCTAA 4320  
DB 4261 TAGGAGAGCCTATGCGAGCTGGAGTGTGTAAGAGGCTCCAGTCTTTGCAATTCCTAA 4320  
QY 4321 AGCCTTTGCTAAAGTTTGGCACTTTTTCATTTTCCATTTTAAAGTAGTTACTA 4380  
DB 4321 AGCCTTTGCTAAAGTTTGGCACTTTTTCATTTTCCATTTTAAAGTAGTTACTA 4380  
QY 4381 AGTTAACTAGTTATTCTGCTTCTGAGTATACGAATTTGGATGCTTAAACCTTATTTTA 4440  
DB 4381 AGTTAACTAGTTATTCTGCTTCTGAGTATACGAATTTGGATGCTTAAACCTTATTTTA 4440  
QY 4441 TAGATGTTATTTAAATATGACAGCAATATCACTCTTTATTGACAACTACCTAAATATGAG 4500  
DB 4441 TAGATGTTATTTAAATATGACAGCAATATCACTCTTTATTGACAACTACCTAAATATGAG 4500  
QY 4501 TTTTATTAATTTAGACTGTAATGGTCTTAAACCACTACTACTGAGAGCTCAATG 4560

[illegible]

D<sub>b</sub> 5581 AAGTTAAATTGTGAAA 5597

## RESULT 2

```

US-09-986-552-4
; Sequence 4, Application US/09986552
; Patent No. US20020150981A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 215089US77DIV
; CURRENT APPLICATION NUMBER: US/09/986,552
; CURRENT FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 5597
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-986-552-4

```

Query Match	100.0%; Score 5597;	DB 9;	Length 5597;
Best Local Similarity	100.0%; Pred. No. 0;		
Matches 5597. Conservative	0. Mismatches	0;	Indels 0;
			Gaps 0;

Qy	1	CGGAGCGAGCGGGCGTCTCGCGCGAGCTGCAATAGCGGCGCCCGGAGCGCTGTGACC	60
Db	1		
Db	1	CGGAGCGAGCGGGCGTCTCGCGCGAGCTGCAATAGCGGCGCCCGGAGCGCTGTGACC	60
Qy	61	TGCGGCGGCGCGCCCGACCGGGGCGCCCTGAAATGGCGGCTCGCTGAGGCGGCGCGCGCGCG	120
Db	61	TGCGGCGGCGCGCCCGACCGGGGCGCCCTGAAATGGCGGCTCGCTGAGGCGGCGCGCGCGCG	120
Qy	121	GGGGCTTCAGGCTCCTCGGGCGTGGCGTGGCGGTGAAGGGGTGATGCTGTTCAGGCTCC	180
Db	121	GGGGCTTCAGGCTCCTCGGGCGTGGCGTGGCGGTGAAGGGGTGATGCTGTTCAGGCTCC	180
Qy	181	TGCAGAGACAAACCTATACCTGTGCCATGGTCCACAGGTATGGGCTCTACGTTCTCTTGG	240
Db	181	TGCAGAGACAAACCTATACCTGTGCCATGGTCCACAGGTATGGGCTCTACGTTCTCTTGG	240
Qy	241	GGTGGTTGTACCAATCGTCTCGGCCCTCCAGTTCGGAGAGGTGGTTCGTGGAATGGAGCC	300
Db	241	GGTGGTTGTACCAATCGTCTCGGCCCTCCAGTTCGGAGAGGTGGTTCGTGGAATGGAGCC	300
Qy	301	GAGATCAATACATGTTTTGTTTGATTCCTATATAGACAAATATGCTGGAAGTCTCTTC	360
Db	301	GAGATCAATACATGTTTTGTTTGATTCCTATATAGACAAATATGCTGGAAGTCTCTTC	360
Qy	361	AGAATCGGCTTGTCTGCCATGCGGATGCAGCTGTGTTACACTGGGTGAATGGCACAG	420
Db	361	AGAATCGGCTTGTCTGCCATGCGGATGCAGCTGTGTTACACTGGGTGAATGGCACAG	420
Qy	421	ATCTTGAACTACTGAAAGAACTACAGCAGGTCTAGAGAACAGATGAGGAGGAGCGAGAAG	480
Db	421	ATCTTGAACTACTGAAAGAACTACAGCAGGTCTAGAGAACAGATGAGGAGGAGCGAGAAG	480
Qy	481	CAATGAGAGAAATCCTTGGGAAAAACAAACGGAACTCTAAGAAGAGTGAAGAAGCAGT	540
Db	481	CAATGAGAGAAATCCTTGGGAAAAACAAACGGAACTCTAAGAAGAGTGAAGAAGCAGT	540
Qy	541	TAGAGTGTTCGTAACACACTGCAATTAAAGTGGCAATGCTTGCTACGACCCAGCCCTGC	600
Db	541	TAGAGTGTTCGTAACACACTGCAATTAAAGTGGCAATGCTTGCTACGACCCAGCCCTGC	600
Qy	601	CAGCCAAACATCACCTGAAGGACGTGCCATCTCTTTATCTCTTCTTTCAATCTGCGAGTG	660
Db	601	CAGCCAAACATCACCTGAAGGACGTGCCATCTCTTTATCTCTTCTTTCAATCTGCGAGTG	660



Db 2821 ATAGTTACTTGGGCTTTTGGCCATGGGAGAAAAAAGTAATTTCCAAAGATCTTCTCGACG 2880  
Qy 2881 AAGAAGAGTCAATGAAGACACAAATGGCATACTTCACTGATAGCAAAAAATCTGGGAGGC 2940  
Db 2881 AAGAAGAGTCAATGAAGACACAAATGGCATACTTCACTGATAGCAAAAAATCTGGGAGGC 2940  
Qy 2941 AACTAAAAGATACATTTGCAGATTTCCCTCAGATATGTAATAAATTTCTAAATAGCAAGT 3000  
Db 2941 AACTAAAAGATACATTTGCAGATTTCCCTCAGATATGTAATAAATTTCTAAATAGCAAGT 3000  
Qy 3001 TTGATTCACATCGCGGAAAGTCCCTGCTCAGATGCTCAGATGATGACCGGATTTGTTA 3060  
Db 3001 TTGATTCACATCGCGGAAAGTCCCTGCTCAGATGCTCAGATGATGACCGGATTTGTTA 3060  
Qy 3061 TGCAGAACTCCAAGATATGTTCCCTGAAGAAATTTGACAGCTGATTTCAAAAGTGC 3120  
Db 3061 TGCAGAACTCCAAGATATGTTCCCTGAAGAAATTTGACAGCTGATTTCAAAAGTGC 3120  
Qy 3121 GCCATTCAGGATATGAGTTTGGCTTCTCTTATTTTATTTATCTCATGAGTGCAGTGC 3180  
Db 3121 GCCATTCAGGATATGAGTTTGGCTTCTCTTATTTTATTTATCTCATGAGTGCAGTGC 3180  
Qy 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATCAGATCAATCTGGTGTCTGT 3240  
Db 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATCAGATCAATCTGGTGTCTGT 3240  
Qy 3241 CTGACAGAGAAATCGGAACATCTGCTACCAAGAAATTCAGAACTGCCGTTAAGTTTCAGG 3300  
Db 3241 CTGACAGAGAAATCGGAACATCTGCTACCAAGAAATTCAGAACTGCCGTTAAGTTTCAGG 3300  
Qy 3301 ATTTGACAGGCTCGGAAACATGCTATATAATTTGCTCAAAAAATGCTTCTGCTGATATCA 3360  
Db 3301 ATTTGACAGGCTCGGAAACATGCTATATAATTTGCTCAAAAAATGCTTCTGCTGATATCA 3360  
Qy 3361 CGCAGCTAAATATATTCACCAATCTCGGAATCTTACTATGATCCAACTGCCACCGG 3420  
Db 3361 CGCAGCTAAATATATTCACCAATCTCGGAATCTTACTATGATCCAACTGCCACCGG 3420  
Qy 3421 TCATCAAAAGTCTAGTAAACAACTGTAACAGTAATCTGACAAATCCACAAAGCATATA 3480  
Db 3421 TCATCAAAAGTCTAGTAAACAACTGTAACAGTAATCTGACAAATCCACAAAGCATATA 3480  
Qy 3481 AGGACAAAAAACAATATAGTTTGAATCATGAGGAGAGAAATCGCTTTTAAATGA 3540  
Db 3481 AGGACAAAAAACAATATAGTTTGAATCATGAGGAGAGAAATCGCTTTTAAATGA 3540  
Qy 3541 TTCGTACCAAGTTTCTCATGTTGTCGAGTTGGATGACATGAAGAAAAACCTTAGGA 3600  
Db 3541 TTCGTACCAAGTTTCTCATGTTGTCGAGTTGGATGACATGAAGAAAAACCTTAGGA 3600  
Qy 3601 AGTTTGTGTTGCTGAATGACAACTTGACCAATATCAAGATGCTCAGACAGTGAAGG 3660  
Db 3601 AGTTTGTGTTGCTGAATGACAACTTGACCAATATCAAGATGCTCAGACAGTGAAGG 3660  
Qy 3661 CTGTTCTCAGGACCTTCTATGAATCCATGTTCCCGTACCTTCCCAATTTGAACTGCCAA 3720  
Db 3661 CTGTTCTCAGGACCTTCTATGAATCCATGTTTCCCGTACCTTCCCAATTTGAACTGCCAA 3720  
Qy 3721 GAGAGTATCGAAACCGTTTCTTATATGATGATGAGCTGACAGAAATGGAGGCTTATCGAG 3780  
Db 3721 GAGAGTATCGAAACCGTTTCTTATATGATGATGAGCTGACAGAAATGGAGGCTTATCGAG 3780  
Qy 3781 ACAAAATTGAAGTTTGGACCCATTTGTPACTAGCAACATTTGATTTACTATATCT 3840  
Db 3781 ACAAAATTGAAGTTTGGACCCATTTGTPACTAGCAACATTTGATTTACTATATCT 3840  
Qy 3841 CATTTTTGCTGAGCAGTTTAAATGACCTTAAGCGGAGATATTTCCAGAGAGGATAC 3900  
Db 3841 CATTTTTGCTGAGCAGTTTAAATGACCTTAAGCGGAGATATTTCCAGAGAGGATAC 3900  
Qy 3901 ACAAAAGAGCTAGTCCCAATCGAATPCAGAGATATGAGATCTTCAATTTGAAAAACCATCTA 3960  
Db 3901 ACAAAAGAGCTAGTCCCAATCGAATPCAGAGATATGAGATCTTCAATTTGAAAAACCATCTA 3960

Qy 3961 CCTCAGCAATTTACTGAGCATTTTAAACTCAGCTTCACAGAGATGCTCTTTGTGATGTGAT 4020  
Db 3961 CCTCAGCAATTTACTGAGCATTTTAAACTCAGCTTCACAGAGATGCTCTTTGTGATGTGAT 4020  
Qy 4021 GCTTAGCAGTTTGGCCCGAAGAAATATCCAGTACCATGCTGTTTTTGTGGCATGAA 4080  
Db 4021 GCTTAGCAGTTTGGCCCGAAGAAATATCCAGTACCATGCTGTTTTTGTGGCATGAA 4080  
Qy 4081 TATAGCCCACTGACTAGGAAATTTATTAACCAACCCACTGAAAACTTGTGTGAGCAGC 4140  
Db 4081 TATAGCCCACTGACTAGGAAATTTATTAACCAACCCACTGAAAACTTGTGTGAGCAGC 4140  
Qy 4141 TCTGAACCTGATTTTACTTTTAAAGAAATTTGCTCATGAGACCTGTCATCTCTTTTATAAAA 4200  
Db 4141 TCTGAACCTGATTTTACTTTTAAAGAAATTTGCTCATGAGACCTGTCATCTCTTTTATAAAA 4200  
Qy 4201 GGCTCACTGACAAAGAGACAGCTGTTTAAATTTCCACAGCAATCATTCAGACATACTTTAT 4260  
Db 4201 GGCTCACTGACAAAGAGACAGCTGTTTAAATTTCCACAGCAATCATTCAGACATACTTTAT 4260  
Qy 4261 TAGGAGAACCTTATGCCAGCTGGGAGTGAATGCTTAAGAGGCTCCAGTCTTTGCAATCCAA 4320  
Db 4261 TAGGAGAACCTTATGCCAGCTGGGAGTGAATGCTTAAGAGGCTCCAGTCTTTGCAATCCAA 4320  
Qy 4321 AGCCCTTTTGTCAAAGTCTTGACACTTTTTTTTCAATTTCCCAATTTTAAAGTAGTACTA 4380  
Db 4321 AGCCCTTTTGTCAAAGTCTTGACACTTTTTTTTCAATTTTCCCAATTTTAAAGTAGTACTA 4380  
Qy 4381 AGTTAACTAGTTATTTCTTGTCTGAGTATTAACGAATTTGGGATGCTTAAACCTATTTTAA 4440  
Db 4381 AGTTAACTAGTTATTTCTTGTCTGAGTATTAACGAATTTGGGATGCTTAAACCTATTTTAA 4440  
Qy 4441 TAGATGTTTATTTAAATAATGACAGCAATATCACCTCTTATTGACCAATACCTAAATATGAG 4500  
Db 4441 TAGATGTTTATTTAAATAATGACAGCAATATCACCTCTTATTGACCAATACCTAAATATGAG 4500  
Qy 4501 TTTTATTAATATTTAAAGACTGTAAGTCTTAAACCACTAACTACTGAAAGAGTCAATG 4560  
Db 4501 TTTTATTAATATTTAAAGACTGTAAGTCTTAAACCACTAACTACTGAAAGAGTCAATG 4560  
Qy 4561 ATTGACATCTGAAATGCTTTGTAATTTGATCTTCAAGCCCTTAAAGATGCTATGATTTCA 4620  
Db 4561 ATTGACATCTGAAATGCTTTGTAATTTGATCTTCAAGCCCTTAAAGATGCTATGATTTCA 4620  
Qy 4621 CGTGCAGGCTCTAAATTTCAACAGGCTTAGATTAATCTACTTACCAGATGTAATATGTTT 4680  
Db 4621 CGTGCAGGCTCTAAATTTCAACAGGCTTAGATTAATCTACTTACCAGATGTAATATGTTT 4680  
Qy 4681 TGGAAATGTACATATTTCAACAGAGTCCCTCATTTTGAAGATGAGTGTGCTGATGGA 4740  
Db 4681 TGGAAATGTACATATTTCAACAGAGTCCCTCATTTTGAAGATGAGTGTGCTGATGGA 4740  
Qy 4741 CTGGCACAATTACAGTGGTGTCTTTTAAATACTCATTTGATATATTTCCAGTAGTATCTCT 4800  
Db 4741 CTGGCACAATTACAGTGGTGTCTTTTAAATACTCATTTGATATATTTCCAGTAGTATCTCT 4800  
Qy 4801 CTCAGTTGGTTTTGATAGAACAGAGGCGCAAACTTTCTTTTGAAGAGCTGTTAGT 4860  
Db 4801 CTCAGTTGGTTTTGATAGAACAGAGGCGCAAACTTTCTTTTGAAGAGCTGTTAGT 4860  
Qy 4861 AAATTTATTTGAGGCGCACTGCTCTTTGTCTATACATTTCTTTGTGTTTGTAGTTGT 4920  
Db 4861 AAATTTATTTGAGGCGCACTGCTCTTTGTCTATACATTTCTTTGTGTTTGTAGTTGT 4920  
Qy 4921 TTTTTCACAAACCTCTTAAATAATGTAATAAACAATGTTTGTGAGCTGTGACATGTAACAAA 4980  
Db 4921 TTTTTCACAAACCTCTTAAATAATGTAATAAACAATGTTTGTGAGCTGTGACATGTAACAAA 4980  
Qy 4981 ACTGCCCCACCGAGATGAGCCCTCAGGCGCATCATTTGCAATCACTGAGAAATATTT 5040  
Db 4981 ACTGCCCCACCGAGATGAGCCCTCAGGCGCATCATTTGCAATCACTGAGAAATATTT 5040





QY 1261 CAATAGTAAACACACGAGATGTTTTGGAATTTGAGCCACTGGCCTACCTTTAGTTTAC 1320  
DB 1261 CAATAGTAAACACACGAGATGTTTTGGAATTTGAGCCACTGGCCTACCTTTAGTTTAC 1320  
QY 1321 CTGCTATTGAAAGTACACATTTCATCGCATCGAAGGGCTGTCCAGAACTTTATTTACCTAA 1380  
DB 1321 CTGCTATTGAAAGTACACATTTCATCGCATCGAAGGGCTGTCCAGAACTTTATTTACCTAA 1380  
QY 1381 ATGATGATGTCATGTTGGGAAAGATGTCTGGCCAGATGATTTTACAGTCACTCCAAAG 1440  
DB 1381 ATGATGATGTCATGTTGGGAAAGATGTCTGGCCAGATGATTTTACAGTCACTCCAAAG 1440  
QY 1441 GCCAGAAGTTTATTTGACATGGGCTGTGCCAACTGTGCCAGGGCTGCCAGGTTTCCT 1500  
DB 1441 GCCAGAAGTTTATTTGACATGGGCTGTGCCAACTGTGCCAGGGCTGCCAGGTTTCCT 1500  
QY 1501 GGATTAAGGATGGCTATTGTGACAAAGCTTGTAAATTTGAGCTTCGATTTGGGATGGTG 1560  
DB 1501 GGATTAAGGATGGCTATTGTGACAAAGCTTGTAAATTTGAGCTTCGATTTGGGATGGTG 1560  
QY 1561 GGGATTGCTCTGGAAACAGTGGAGGAGTCCCTATATTGACAGGAGTGGAGTACTGGGA 1620  
DB 1561 GGGATTGCTCTGGAAACAGTGGAGGAGTCCCTATATTGACAGGAGTGGAGTACTGGGA 1620  
QY 1621 GTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTCTTACTGTA 1680  
DB 1621 GTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTCTTACTGTA 1680  
QY 1681 ATCAGGATGTCGAAATTCCTGGCTCGCTGATATAGTTCTGTGACCAAGCATGCAATGTCT 1740  
DB 1681 ATCAGGATGTCGAAATTCCTGGCTCGCTGATATAGTTCTGTGACCAAGCATGCAATGTCT 1740  
QY 1741 TGTCTGTGGGTTGATCTGGCAGCTGTGGCAGAGTCAATTTTCATGAATTTGTATAAG 1800  
DB 1741 TGTCTGTGGGTTGATCTGGCAGCTGTGGCAGAGTCAATTTTCATGAATTTGTATAAG 1800  
QY 1801 TGATCTCTTCCAAACAGACTCACTATATATTTCGAAAGTGAATGCTGCTGCTATT 1860  
DB 1801 TGATCTCTTCCAAACAGACTCACTATATATTTCGAAAGTGAATGCTGCTGCTATT 1860  
QY 1861 TCAGCTTTGCAAGATGACCAAAAGAGAGTGTGAAGGTGCCTATAGTGACATCCAAATA 1920  
DB 1861 TCAGCTTTGCAAGATGACCAAAAGAGAGTGTGAAGGTGCCTATAGTGACATCCAAATA 1920  
QY 1921 TTCGACATGCTTCTATTCCCAACAGTGGAAAAACATCCACCTCATATGTCAGAGTGGAA 1980  
DB 1921 TTCGACATGCTTCTATTCCCAACAGTGGAAAAACATCCACCTCATATGTCAGAGTGGAA 1980  
QY 1981 TGAATGCCACCAATACATTTTAACTCTCAGCTTTCAAATACAAACGATGAGAGTTCA 2040  
DB 1981 TGAATGCCACCAATACATTTTAACTCTCAGCTTTCAAATACAAACGATGAGAGTTCA 2040  
QY 2041 AAATCAGATTAACAGTGGAGTGGACACAAAGGAGGACCAAACTGAATTTCTACGGCCC 2100  
DB 2041 AAATCAGATTAACAGTGGAGTGGACACAAAGGAGGACCAAACTGAATTTCTACGGCCC 2100  
QY 2101 AGAAGGTTACGAAATTTAGTTAGTCCCTAAACATCTCTTCAGAGCGGAATCCTTT 2160  
DB 2101 AGAAGGTTACGAAATTTAGTTAGTCCCTAAACATCTCTTCAGAGCGGAATCCTTT 2160  
QY 2161 TTGAGGATATCCCAAGAAAAACGCTTCCCGAAGTTTAAAGACATGATGTTAACTCAA 2220  
DB 2161 TTGAGGATATCCCAAGAAAAACGCTTCCCGAAGTTTAAAGACATGATGTTAACTCAA 2220  
QY 2221 CAAAGAGAGCCAGGAGAGTGAATAATTCCTCGGTAAATATTTCATCTCCTCCAAAG 2280  
DB 2221 CAAAGAGAGCCAGGAGAGTGAATAATTCCTCGGTAAATATTTCATCTCCTCCAAAG 2280  
QY 2281 ACGCCAGTTGATGCTCAATACCTTGGATTTGCAACTGGAACATGGAGACATCACTTTGA 2340  
DB 2281 ACGCCAGTTGATGCTCAATACCTTGGATTTGCAACTGGAACATGGAGACATCACTTTGA 2340

QY 2341 AAGGATACAAATTTGTCCAAGTCAGCCTTGCTGAGATCAATTTCTGATGAACCTCACAGCATG 2400  
DB 2341 AAGGATACAAATTTGTCCAAGTCAGCCTTGCTGAGATCAATTTCTGATGAACCTCACAGCATG 2400  
QY 2401 CTAAATAAAAATCAAGCTATATAACAGATGAACAAATGACAGTTTGGTGGCTCCAC 2460  
DB 2401 CTAAATAAAAATCAAGCTATATAACAGATGAACAAATGACAGTTTGGTGGCTCCAC 2460  
QY 2461 AGGAAAAACAGGTTTCATATAAGCATTTGCCAAAACAGCTTAGGAGTGTCTGAAAGATTGC 2520  
DB 2461 AGGAAAAACAGGTTTCATATAAGCATTTGCCAAAACAGCTTAGGAGTGTCTGAAAGATTGC 2520  
QY 2521 AGAGGTTGATCTTTTCTGCAAGTGTAAAGTGAATGGTCACTGACAGGGTCAAGATC 2580  
DB 2521 AGAGGTTGATCTTTTCTGCAAGTGTAAAGTGAATGGTCACTGACAGGGTCAAGATC 2580  
QY 2581 CACCCCTGGACTTGGAGACCAAGCAAGATTTAGAGTGGAACTCACACCCAAAAAACCA 2640  
DB 2581 CACCCCTGGACTTGGAGACCAAGCAAGATTTAGAGTGGAACTCACACCCAAAAAACCA 2640  
QY 2641 TAGGCGGAAATGTGACAAAAGAAAAAGCCCCCATCTCTGATTTGTTCCACTGGAAAGCCAGA 2700  
DB 2641 TAGGCGGAAATGTGACAAAAGAAAAAGCCCCCATCTCTGATTTGTTCCACTGGAAAGCCAGA 2700  
QY 2701 TGACAAAAGAAAAATCACAGGAAAGAAAAAGAAACAGTAGATGAGGAAATG 2760  
DB 2701 TGACAAAAGAAAAATCACAGGAAAGAAAAAGAAACAGTAGATGAGGAAATG 2760  
QY 2761 CTGAAATACATATAGGCGTTACTGAAAGTGTACTTGGAAAGAAAGCTGCAGCATTAACAG 2820  
DB 2761 CTGAAATACATATAGGCGTTACTGAAAGTGTACTTGGAAAGAAAGCTGCAGCATTAACAG 2820  
QY 2821 ATAGTTACTTGGGCTTTTGGCATGGGAAAGAAAAAGTATTTCCAGATCTTCTCGAGC 2880  
DB 2821 ATAGTTACTTGGGCTTTTGGCATGGGAAAGAAAAAGTATTTCCAGATCTTCTCGAGC 2880  
QY 2881 AAGAAGAGTCATTGAAAGACACAAATTTGGCATCTTCACTGATAGCAAAATACTGGGAGGC 2940  
DB 2881 AAGAAGAGTCATTGAAAGACACAAATTTGGCATCTTCACTGATAGCAAAATACTGGGAGGC 2940  
QY 2941 AAATAAAGATACATTTGCAGATTCCTCAGATATGTAATAAATTTCTAAATAGCAAGT 3000  
DB 2941 AAATAAAGATACATTTGCAGATTCCTCAGATATGTAATAAATTTCTAAATAGCAAGT 3000  
QY 3001 TTGGATTCATATCGGAAAGTCCCTGCTCATGSCCTCATGATTTGACCGGATTTGTTA 3060  
DB 3001 TTGGATTCATATCGGAAAGTCCCTGCTCATGSCCTCATGATTTGACCGGATTTGTTA 3060  
QY 3061 TGCAAGAACTGCAAGATATGTTCCCTGAAGATTTGCAAGACGTCATTTCAAAAGTGC 3120  
DB 3061 TGCAAGAACTGCAAGATATGTTCCCTGAAGATTTGCAAGACGTCATTTCAAAAGTGC 3120  
QY 3121 GGCATTTGAGGATATGAGTTTGGCTTCTCTTATTTTATCTCATGATGAGTGCAGTGC 3180  
DB 3121 GGCATTTGAGGATATGAGTTTGGCTTCTCTTATTTTATCTCATGATGAGTGCAGTGC 3180  
QY 3181 AGCCACTGAATATATCTCAAAGTCTTTGATGAAGTTGATACAGATCAATCTGCTGCTTGT 3240  
DB 3181 AGCCACTGAATATATCTCAAAGTCTTTGATGAAGTTGATACAGATCAATCTGCTGCTTGT 3240  
QY 3241 CTGACAGAGAAATCCGAAACACATGGCTACAGAAATTCAGAACTGCGGTAACTTTGCGAG 3300  
DB 3241 CTGACAGAGAAATCCGAAACACATGGCTACAGAAATTCAGAACTGCGGTAACTTTGCGAG 3300  
QY 3301 ATTTGACAGCTGGAACACATGCTTAATAATTTGCTCAAAATGCTTCTGCTGATATCA 3360  
DB 3301 ATTTGACAGCTGGAACACATGCTTAATAATTTGCTCAAAATGCTTCTGCTGATATCA 3360  
QY 3361 CGCAGCTAAATATTTCCAACTCAGGAATCTTACTGATCCCAACCTGCCACCGG 3420  
DB 3361 CGCAGCTAAATATTTCCAACTCAGGAATCTTACTGATCCCAACCTGCCACCGG 3420  
QY 3421 TCACTAAAAGTCTAGTAAACAACTGTAAACAGTAACTGACAAAATCCAAAGCATATATA 3480



Db 3421 TCACTAAAAGCTAGTAAACAACTGTAAACCAAGTAACTGACAAAATCCACAAAGCATATA 3480  
Qy 3481 AGGACAAAACAAATATAGGTTTGAATCATGGGAGAGAAAGAAATCCCTTTTAAAATGA 3540  
Db 3481 AGGACAAAACAAATATAGGTTTGAATCATGGGAGAGAAAGAAATCCCTTTTAAAATGA 3540  
Qy 3541 TTCGTACCAACGTTTCTCATGTGGTTGGCCAGTGGATGACATTAAGAAAACCCCTAGGA 3600  
Db 3541 TTCGTACCAACGTTTCTCATGTGGTTGGCCAGTGGATGACATTAAGAAAACCCCTAGGA 3600  
Qy 3601 AGTTTGTTCCTGCTGAATGACAAATGACCAATGACCAATCATTAAGATGCTCAGACAGTGAAG 3660  
Db 3601 AGTTTGTTCCTGCTGAATGACCAATGACCAATCATTAAGATGCTCAGACAGTGAAG 3660  
Qy 3661 CTGTTCTCAGGACCTTATGAATCCATGTTCCCATACCTTCCCAATTTGAATGACCA 3720  
Db 3661 CTGTTCTCAGGACCTTATGAATCCATGTTCCCATACCTTCCCAATTTGAATGACCA 3720  
Qy 3721 GAGAGTATCGAAACCGTTTCCCTTATGATGATGATGATGATGATGATGATGATGATGAT 3780  
Db 3721 GAGAGTATCGAAACCGTTTCCCTTATGATGATGATGATGATGATGATGATGATGATGAT 3780  
Qy 3781 ACAAAATGAGTTTGGACCCATGTTGATGATGATGATGATGATGATGATGATGATGAT 3840  
Db 3781 ACAAAATGAGTTTGGACCCATGTTGATGATGATGATGATGATGATGATGATGATGAT 3840  
Qy 3841 CATTTTCTGAGCAGTTAATGACCTTACGAGATGATGATGATGATGATGATGATGATGAT 3900  
Db 3841 CATTTTCTGAGCAGTTAATGACCTTACGAGATGATGATGATGATGATGATGATGATGAT 3900  
Qy 3901 ACAAAAGAGCTAGTCCCAATCGAATCAGATATAGAGATCTTCAATTTGAAACCCATCTA 3960  
Db 3901 ACAAAAGAGCTAGTCCCAATCGAATCAGATATAGAGATCTTCAATTTGAAACCCATCTA 3960  
Qy 3961 CCTCAGATTTACTGAGCATTTTAAACTCAGCTTACAGAGATGCTTTGTGATGAT 4020  
Db 3961 CCTCAGATTTACTGAGCATTTTAAACTCAGCTTACAGAGATGCTTTGTGATGAT 4020  
Qy 4021 GCTTACGCTTTGGCCGAGAGAAATATCCAGTACCATGCTGTTTGTGGCATGAA 4080  
Db 4021 GCTTACGCTTTGGCCGAGAGAAATATCCAGTACCATGCTGTTTGTGGCATGAA 4080  
Qy 4081 TATAGCCCACTCAGTAGGAATTTAAACCAACCACTGAAACCTTGTGTCGAGCAGC 4140  
Db 4081 TATAGCCCACTCAGTAGGAATTTAAACCAACCACTGAAACCTTGTGTCGAGCAGC 4140  
Qy 4141 TCTGACTGATTTTACTTTTAAAGAAATTTGCTCATGGACCTGCTATCTTTTATATAAA 4200  
Db 4141 TCTGACTGATTTTACTTTTAAAGAAATTTGCTCATGGACCTGCTATCTTTTATATAAA 4200  
Qy 4201 GGCTCACTGACAGAGACAGCTGTTAAATTTCCACAGCAATCATTCGAGACTAATTTAT 4260  
Db 4201 GGCTCACTGACAGAGACAGCTGTTAAATTTCCACAGCAATCATTCGAGACTAATTTAT 4260  
Qy 4261 TAGGAGAGCCCTATGCCAGCTGGAGTGATGCTTAAGAGGCTCAGCTTTGTCATCCAA 4320  
Db 4261 TAGGAGAGCCCTATGCCAGCTGGAGTGATGCTTAAGAGGCTCAGCTTTGTCATCCAA 4320  
Qy 4321 AGCCTTTTCTAAAGTTTGGACCTTTTTCATTTCCATTTTCAATTTTAAAGTAGTACTA 4380  
Db 4321 AGCCTTTTCTAAAGTTTGGACCTTTTTCATTTCCATTTTCAATTTTAAAGTAGTACTA 4380  
Qy 4381 AGTTAACTAGTTATTTCTGCTCTGAGTATACGAATTTGGGATGCTAAACCTATTTTAA 4440  
Db 4381 AGTTAACTAGTTATTTCTGCTCTGAGTATACGAATTTGGGATGCTAAACCTATTTTAA 4440  
Qy 4441 TAGATGTTATTTAAATTAATGACAGCAATACCTCTTATGCAATACCTAAATTTATGAG 4500  
Db 4441 TAGATGTTATTTAAATTAATGACAGCAATACCTCTTATGCAATACCTAAATTTATGAG 4500  
Qy 4501 TTTTATTAATTTTAAAGCTGTAATGGTCTTTAAACCACTAACTACTGAGAGCTCAATG 4560

Db 4501 TTTTATTAATTTTAAAGCTGTAATGGTCTTTAAACCACTAACTACTGAGAGCTCAATG 4560  
Qy 4561 ATTGACATCTGAAATGCTTTGTAATTTATTTGACTTCAGCCCTTAAGAAATGCTATGATTTCA 4620  
Db 4561 ATTGACATCTGAAATGCTTTGTAATTTATTTGACTTCAGCCCTTAAGAAATGCTATGATTTCA 4620  
Qy 4621 CGTGCAGGCTTAATTTCAACAGGCTAGAGTTAGTACTTACACAGATGTAATTAATGTTT 4680  
Db 4621 CGTGCAGGCTTAATTTCAACAGGCTAGAGTTAGTACTTACACAGATGTAATTAATGTTT 4680  
Qy 4681 TGGAAATGTAATTTCAACAGAGTGCTCATTTTGAAGATGAGTAGTGTGATGGCA 4740  
Db 4681 TGGAAATGTAATTTCAACAGAGTGCTCATTTTGAAGATGAGTAGTGTGATGGCA 4740  
Qy 4741 CTGGCAGATTTACAGTGGTCTTCTTTTAACTATCATTTGATTTCCAGTAGTATCTCT 4800  
Db 4741 CTGGCAGATTTACAGTGGTCTTCTTTTAACTATCATTTGATTTCCAGTAGTATCTCT 4800  
Qy 4801 CTGAGTTGGTTTGTATAGAAACAGAGCCAGCAAACTTTCTTTGTAAGAGCTGGTTAGT 4860  
Db 4801 CTGAGTTGGTTTGTATAGAAACAGAGCCAGCAAACTTTCTTTGTAAGAGCTGGTTAGT 4860  
Qy 4861 AAATTAATGAGCCAGCTGCTGTTTGTGATACATCTTTCTTCTGCTGTTGTTAGTTGT 4920  
Db 4861 AAATTAATGAGCCAGCTGCTGTTTGTGATACATCTTTCTTCTGCTGTTGTTAGTTGT 4920  
Qy 4921 TTTTCTTCAAAACCAACCTCTTAAATGTAAGAAACCACTTTCTTTGTAAGAGCTGGTTAGT 4980  
Db 4921 TTTTCTTCAAAACCAACCTCTTAAATGTAAGAAACCACTTTCTTTGTAAGAGCTGGTTAGT 4980  
Qy 4981 ACTGCCACAGCCAGATGAGCCCTCAGCCATCATTTGTCCTGCTGCTGCTGCTGCTGCT 5040  
Db 4981 ACTGCCACAGCCAGATGAGCCCTCAGCCATCATTTGTCCTGCTGCTGCTGCTGCTGCT 5040  
Qy 5041 TTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 5100  
Db 5041 TTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 5100  
Qy 5101 TGCAGTGGCCCAATCTCAGCTCACTGCAACCTCCGCTCCCGGGTTCAAGCAGTTCTGTC 5160  
Db 5101 TGCAGTGGCCCAATCTCAGCTCACTGCAACCTCCGCTCCCGGGTTCAAGCAGTTCTGTC 5160  
Qy 5161 TCAGCTTCTGAGTAGTGGGACTACAGGTGATGCAACACACCTGCTGCTGCTGCTGCTGCT 5220  
Db 5161 TCAGCTTCTGAGTAGTGGGACTACAGGTGATGCAACACACCTGCTGCTGCTGCTGCTGCT 5220  
Qy 5221 TTTTGTAGTAGAGAGGGGGTTCCACCATATTGCTCAGGCTTATCTTGAACCTCCTGACCTC 5280  
Db 5221 TTTTGTAGTAGAGAGGGGGTTCCACCATATTGCTCAGGCTTATCTTGAACCTCCTGACCTC 5280  
Qy 5281 AGTGTATCCACTGCTGCTCCCAAGTGTGAGATTAAGGCAATAGCCAGTGCAC 5340  
Db 5281 AGTGTATCCACTGCTGCTCCCAAGTGTGAGATTAAGGCAATAGCCAGTGCAC 5340  
Qy 5341 CCAGCCGAGAAATAGTATTTTATGTTGTTTAAACCTTGGCGCTAGCCATTTTATG 5400  
Db 5341 CCAGCCGAGAAATAGTATTTTATGTTGTTTAAACCTTGGCGCTAGCCATTTTATG 5400  
Qy 5401 TCATATATCAATGGAATTTGTAAGAGAGCAGATTTCCATGAGTAACTCTGACAGGATTTTATG 5460  
Db 5401 TCATATATCAATGGAATTTGTAAGAGAGCAGATTTCCATGAGTAACTCTGACAGGATTTTATG 5460  
Qy 5461 ATCATGATCTCAACATATTTCTCCCAATGAGTACATCTTTTGTACAAAGACTGAA 5520  
Db 5461 ATCATGATCTCAACATATTTCTCCCAATGAGTACATCTTTTGTACAAAGACTGAA 5520  
Qy 5521 ATGTAAATACTGTTGTTGCTGCTGTAAGAGTTGTTGTTTCAAAACCTGAAATCTCAAAA 5580  
Db 5521 ATGTAAATACTGTTGTTGCTGCTGTAAGAGTTGTTGTTTCAAAACCTGAAATCTCAAAA 5580  
Qy 5581 AAGTTAAATTTGAAAA 5597  
Db 5581 AAGTTAAATTTGAAAA 5597



Db 1801 TGATCCTTCTCCAAACAGACTCACTATATATATCCAAAAGGTGAATGCCCTGCTTATT 1860  
Qy 1861 TCAGCTTTGCAGAGTAGCCAAAGAGAGAGTGAAGTGCCTATAGTACAGATCAATCCAAATA 1920  
Db 1861 TCAGCTTTGCAGAGTAGCCAAAGAGAGAGTGAAGTGCCTATAGTACAGATCAATCCAAATA 1920  
Qy 1921 TTCGACATGCTTCTATTCGCAACAGTGGAAACCATCCACTCATATATGACACAGTGGAA 1980  
Db 1921 TTCGACATGCTTCTATTCGCAACAGTGGAAACCATCCACTCATATATGACACAGTGGAA 1980  
Qy 1981 TGAATGCCACCAACAATATTTAATCTCAGCTTTCAAAATACAAACGATGAAGAGTTCA 2040  
Db 1981 TGAATGCCACCAACAATATTTAATCTCAGCTTTCAAAATACAAACGATGAAGAGTTCA 2040  
Qy 2041 AAATCGACATACAGTGGAGTGGACACAAGGGGAGGACCAAACTGAATTTCTAGGGCCC 2100  
Db 2041 AAATCGACATACAGTGGAGTGGACACAAGGGGAGGACCAAACTGAATTTCTAGGGCCC 2100  
Qy 2101 AGAAGGGTTACGAAATTTAGTTAGTCCCATAACTCTTCCAGAGGGCGAAATCCCTTT 2160  
Db 2101 AGAAGGGTTACGAAATTTAGTTAGTCCCATAACTCTTCCAGAGGGCGAAATCCCTTT 2160  
Qy 2161 TTGAGGATATCCCAAGAAACGCTTCCCGAAGTTTAAGAGACATGATGTTAACTCAA 2220  
Db 2161 TTGAGGATATCCCAAGAAACGCTTCCCGAAGTTTAAGAGACATGATGTTAACTCAA 2220  
Qy 2221 CAAGGAGAGCCAGGAGAGGTGAAATTCCTCGTAAATTTCACTCTCTCCCAAG 2280  
Db 2221 CAAGGAGAGCCAGGAGAGGTGAAATTCCTCGTAAATTTCACTCTCTCCCAAG 2280  
Qy 2281 AGCCGAGTGTAGTCTCAATACCTTTGGATTTGCAATCGGAACATGAGATCACTTTGA 2340  
Db 2281 AGCCGAGTGTAGTCTCAATACCTTTGGATTTGCAATCGGAACATGAGATCACTTTGA 2340  
Qy 2341 AAGGATACAAATTTGTCAGAGTCAAGCTTGTGAGATCACTTCTGATGAATTCACAGCATG 2400  
Db 2341 AAGGATACAAATTTGTCAGAGTCAAGCTTGTGAGATCACTTCTGATGAATTCACAGCATG 2400  
Qy 2401 CTAAATAAAAATCAAGCTATAATAACAGATGAATGAACAAATGACAGTTTGGTGGCTCCAC 2460  
Db 2401 CTAAATAAAAATCAAGCTATAATAACAGATGAATGAACAAATGACAGTTTGGTGGCTCCAC 2460  
Qy 2461 AGGAAACAGGTTCTATAAAGCATCTTGCCAAACAGCTTAGGAGTGTCTGAAAGATTGC 2520  
Db 2461 AGGAAACAGGTTCTATAAAGCATCTTGCCAAACAGCTTAGGAGTGTCTGAAAGATTGC 2520  
Qy 2521 AGAGTTGACTTTTCTCGAGTGAAGTGAATGCTCATGACAGGAGTCAAGATC 2580  
Db 2521 AGAGTTGACTTTTCTCGAGTGAAGTGAATGCTCATGACAGGAGTCAAGATC 2580  
Qy 2581 CACCCCTGACCTTGAGACCAAGCAGATTTAGAGTGGAAACTCAGACCCCAAAACCA 2640  
Db 2581 CACCCCTGACCTTGAGACCAAGCAGATTTAGAGTGGAAACTCAGACCCCAAAACCA 2640  
Qy 2641 TAGGGGAAATGTGACAAAGAAAGCCCCCATCTCTGATTTCCACTGGAAGGCCAGA 2700  
Db 2641 TAGGGGAAATGTGACAAAGAAAGCCCCCATCTCTGATTTCCACTGGAAGGCCAGA 2700  
Qy 2701 TGACAAAAGAAAGAAATTCACAGGAAAGAAAGAGAACAGTGAATGGAGGAAATG 2760  
Db 2701 TGACAAAAGAAAGAAATTCACAGGAAAGAAAGAGAACAGTGAATGGAGGAAATG 2760  
Qy 2761 CTGAAATCAGATAGCGTTTACTGAGTGTACTTGGAGAAAGCTGAGCAATTACAG 2820  
Db 2761 CTGAAATCAGATAGCGTTTACTGAGTGTACTTGGAGAAAGCTGAGCAATTACAG 2820  
Qy 2821 ATAGTTTACCTTGGCTTTTGGCATGGGAAAGAAAGTATTTCCAAGATCTTCTCGAGC 2880  
Db 2821 ATAGTTTACCTTGGCTTTTGGCATGGGAAAGAAAGTATTTCCAAGATCTTCTCGAGC 2880  
Qy 2881 AAGAGAGTCATTGAAGACAAATTTGGCATCTTCTGATAGCAAAATACTGGGAGGC 2940

Db 2881 AAGAGAGTCATTGAAGACACAATTTGGCATCTTCTGATAGCAAAATACTGGGAGGC 2940  
Qy 2941 AACTAAAGATACATTTGAGATTTCCCTCAGATATGTAATAATAATAATTTCTAAATAGCAAGT 3000  
Db 2941 AACTAAAGATACATTTGAGATTTCCCTCAGATATGTAATAATAATAATTTCTAAATAGCAAGT 3000  
Qy 3001 TTGATTTCAATCGCGGAAAGTCCCTGCTCACATGCTCACATGATTTGACCGGATTTGTTA 3060  
Db 3001 TTGATTTCAATCGCGGAAAGTCCCTGCTCACATGCTCACATGATTTGACCGGATTTGTTA 3060  
Qy 3061 TGCAGAACTGCAAGATATGTTCCCTGAAGATTTGACAAAGCTCATTTCAAAAGTGC 3120  
Db 3061 TGCAGAACTGCAAGATATGTTCCCTGAAGATTTGACAAAGCTCATTTCAAAAGTGC 3120  
Qy 3121 GCCATTTGAGGATATGCGAGTTTGCCTTCTCTTATTTTATTTATCTCATGAGTGCAGTGC 3180  
Db 3121 GCCATTTGAGGATATGCGAGTTTGCCTTCTCTTATTTTATTTATCTCATGAGTGCAGTGC 3180  
Qy 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTGTATACAGATCAATCTGTTGTTCTTGT 3240  
Db 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTGTATACAGATCAATCTGTTGTTCTTGT 3240  
Qy 3241 CTGACAGAGAAATCCGAAACACTGCTTACCAGAAATTCACGAACTGCCGTTAAAGTTTGCAGG 3300  
Db 3241 CTGACAGAGAAATCCGAAACACTGCTTACCAGAAATTCACGAACTGCCGTTAAAGTTTGCAGG 3300  
Qy 3301 ATTTGACAGTCTGGAACACATGCTAATAATTTGCTCAAAAATTTGCTTCTGCTGATATCA 3360  
Db 3301 ATTTGACAGTCTGGAACACATGCTAATAATTTGCTCAAAAATTTGCTTCTGCTGATATCA 3360  
Qy 3361 CGCAGCTTAATATATTTCCACCAACTCAGGAATCTTACTATGATCCCACTGCCACCGG 3420  
Db 3361 CGCAGCTTAATATATTTCCACCAACTCAGGAATCTTACTATGATCCCACTGCCACCGG 3420  
Qy 3421 TCCTAAAAAGCTAGTAACAAACTGTAAACCACTGTAACCACTGACAAAATCCCAAGCATATA 3480  
Db 3421 TCCTAAAAAGCTAGTAACAAACTGTAAACCACTGTAACCACTGACAAAATCCCAAGCATATA 3480  
Qy 3481 AGGACAAAACAAATATAGTTTGAATCATGGGAGAAAGAAATCGCTTTTAAATGA 3540  
Db 3481 AGGACAAAACAAATATAGTTTGAATCATGGGAGAAAGAAATCGCTTTTAAATGA 3540  
Qy 3541 TTGCTACCAACGTTTCTCATGTGCTTGGCCAGTTGGATGACATAAGAAAACCTTAGGA 3600  
Db 3541 TTGCTACCAACGTTTCTCATGTGCTTGGCCAGTTGGATGACATAAGAAAACCTTAGGA 3600  
Qy 3601 AGTTTGTTCCTGAAATGACAACTTGAACAATCATAAAGATGCTCAGACAGTGAAGG 3660  
Db 3601 AGTTTGTTCCTGAAATGACAACTTGAACAATCATAAAGATGCTCAGACAGTGAAGG 3660  
Qy 3661 CTGTTCTCAGGACTTCTATGAAATCCATGTTCCCATACCTTCCCAATTTGAACCTGCCAA 3720  
Db 3661 CTGTTCTCAGGACTTCTATGAAATCCATGTTCCCATACCTTCCCAATTTGAACCTGCCAA 3720  
Qy 3721 GAGAGTATCGAAACCGTTTCTCATGATGAGCTGCAAGAAATGGAGGCTTATCGAG 3780  
Db 3721 GAGAGTATCGAAACCGTTTCTCATGATGAGCTGCAAGAAATGGAGGCTTATCGAG 3780  
Qy 3781 ACAATTTGAAGTTTGGACCCATGTTACTAGCAACATGATTATGTTTACTATATCT 3840  
Db 3781 ACAATTTGAAGTTTGGACCCATGTTACTAGCAACATGATTATGTTTACTATATCT 3840  
Qy 3841 CATTTTGTGCTGAGCAGTTAAATTTGCATTTAAGCGGAGATATTTCCCAAGAGGAGATAC 3900  
Db 3841 CATTTTGTGCTGAGCAGTTAAATTTGCATTTAAGCGGAGATATTTCCCAAGAGGAGATAC 3900  
Qy 3901 ACAAGAGCTAGTCCCAATCGAATCAGAGTATAGAAAGATCTTTCATTTGAAACCATCTA 3960  
Db 3901 ACAAGAGCTAGTCCCAATCGAATCAGAGTATAGAAAGATCTTTCATTTGAAACCATCTA 3960  
Qy 3961 CCTCAGCATTTACTGACATTTTAAACTCAGCTTCCAGAGATGCTTTTGTGATGTGAT 4020  
Db 3961 CCTCAGCATTTACTGACATTTTAAACTCAGCTTCCAGAGATGCTTTTGTGATGTGAT 4020

QY	4021	GCTTAGCAGTTTGGCCCCGAAGAAGAAATAATCCAGTACCATGCTGTTTTGTGTCATGAA	4080
DB	4021	GCTTAGCAGTTTGGCCCCGAAGAAGAAATAATCCAGTACCATGCTGTTTTGTGTCATGAA	4080
QY	4081	TATAGCCCACTGACTAGGAAATTATTTAAACCAACCCTGAAACCTGCTGTGTCGAGCAGC	4140
DB	4081	TATAGCCCACTGACTAGGAAATTATTTAAACCAACCCTGAAACCTGCTGTGTCGAGCAGC	4140
QY	4141	TCTGAACCTGATTTTACTTTTTAAAGAAATTGCTCATGCACTGTCATCCTTTTTATAAAAA	4200
DB	4141	TCTGAACCTGATTTTACTTTTTAAAGAAATTGCTCATGCACTGTCATCCTTTTTATAAAAA	4200
QY	4201	GGCTCACTGCAAGACAGACAGCTGTTAATTTCCCAACAGCAATCATGTCAGACTACTTTTAT	4260
DB	4201	GGCTCACTGCAAGACAGACAGCTGTTAATTTCCCAACAGCAATCATGTCAGACTACTTTTAT	4260
QY	4261	TAGGAGAAGCCTATGCCAGCTGGGAGTGATTGCTPAAGAGGCTCCAGTCTTTGCAATCCAA	4320
DB	4261	TAGGAGAAGCCTATGCCAGCTGGGAGTGATTGCTPAAGAGGCTCCAGTCTTTGCAATCCAA	4320
QY	4321	AGCCTTTTGCTAARAGTTTTCGACCTTTTTTTTTTTTCAATTTCCCATTTTAAAGTAGTTACTA	4380
DB	4321	AGCCTTTTGCTAARAGTTTTCGACCTTTTTTTTTTTTCAATTTCCCATTTTAAAGTAGTTACTA	4380
QY	4381	AGTTAACTAGTATATCTTTGCTCTCGAGTATAACGAAATGGGATGCTCTAAACCTATTTTTA	4440
DB	4381	AGTTAACTAGTATATCTTTGCTCTCGAGTATAACGAAATGGGATGCTCTAAACCTATTTTTA	4440
QY	4441	TAGATCTTATTTAAATAATGCAGCAATATCACTCTTTATTGACAAATACCTAAATTTATGAG	4500
DB	4441	TAGATCTTATTTAAATAATGCAGCAATATCACTCTTTATTGACAAATACCTAAATTTATGAG	4500
QY	4501	TTTTTATTAATTTTAAAGACTGTAATGGTCTTTAAACCACCTAACTACTGAAAGAGCTCAATG	4560
DB	4501	TTTTTATTAATTTTAAAGACTGTAATGGTCTTTAAACCACCTAACTACTGAAAGAGCTCAATG	4560
QY	4561	ATTGACATCTGAATAGCTTTGTATTTATTTGACTTCAGCCCCCTAAGANTGCTATGATTTCA	4620
DB	4561	ATTGACATCTGAATAGCTTTGTATTTATTTGACTTCAGCCCCCTAAGANTGCTATGATTTCA	4620
QY	4621	CGTGCAGGTCCTAATTTCAAACAGGCTAGAGTTAGTACTACTTACAGATGTAATATGTTTT	4680
DB	4621	CGTGCAGGTCCTAATTTCAAACAGGCTAGAGTTAGTACTACTTACAGATGTAATATGTTTT	4680
QY	4681	TGGAATATGTACATATTCAAACAGAGTGCCTCATTTTGAAGATGAGTAGTCTGATGGCA	4740
DB	4681	TGGAATATGTACATATTCAAACAGAGTGCCTCATTTTGAAGATGAGTAGTCTGATGGCA	4740
QY	4741	CTGGCAATATACAGTGGTGCTCTGTTTAAACTCATTGGTATATTCAGTAGCTATCTCT	4800
DB	4741	CTGGCAATATACAGTGGTGCTCTGTTTAAACTCATTGGTATATTCAGTAGCTATCTCT	4800
QY	4801	CTCAGTTGGTTTTGATAGAACAGAGGCCAGCAAACTTTCTTTGTAAGAGCTGGTTAGT	4860
DB	4801	CTCAGTTGGTTTTGATAGAACAGAGGCCAGCAAACTTTCTTTGTAAGAGCTGGTTAGT	4860
QY	4861	AAATTAATGACGGCCACTGTGTCTTTGTGCATACATTTCTTCTGCTCTGTTAGTTTGT	4920
DB	4861	AAATTAATGACGGCCACTGTGTCTTTGTGCATACATTTCTTCTGCTCTGTTAGTTTGT	4920
QY	4921	TTTTTTTTCAAACAAACCTCTAAATAATGTAAAAACCATGTTTAGCTTGCAGCTGTACAAA	4980
DB	4921	TTTTTTTTCAAACAAACCTCTAAATAATGTAAAAACCATGTTTAGCTTGCAGCTGTACAAA	4980
QY	4981	ACTGCCACACAGCCAGATGTGACCCCTCAGGCCATCATTTGCCAATCACTCAGAAATATTT	5040
DB	4981	ACTGCCACACAGCCAGATGTGACCCCTCAGGCCATCATTTGCCAATCACTCAGAAATATTT	5040
QY	5041	TTGTTGTTGTTGTTGTTGTTTGTGAGACAGAGTCTCTCTGTTGCCAGGCTGGAG	5100
DB	5041	TTGTTGTTGTTGTTGTTGTTTGTGAGACAGAGTCTCTCTGTTGCCAGGCTGGAG	5100

## RESULTS

RESULT 5  
US-10-023-890-3

US-10-023-890-3

; Sequence 3, Application US/1002

; Publication No. US20

; GENERAL INFORMATION:

APPLICANT: CANFIELD, William

**TITLE OF INVENTION: METHOD OF PRODUCING**

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

FILE OF INVENTION: MANTALIAN CELLS  
FILE REFERENCE: 203510US77

```

; FILE REFERENCE: 203510US//
; CURRENT APPLICATION NUMBER: IIS/1

```

	CURRENT APPLICATION NUMBER	CURRENT FILING DATE	CURRENT PRIORITY DATE
;			

; CURRENT FILING DATE: 2001-12-

; NUMBER OF SEQ ID NOS: 21

; SOFTWARE: Pat

; SEQ ID NO 3

; LENGTH: 5597

```
; LENGTH: 333
: TYPE: PNA
```

Query Match 100.0%: Score 5597: DB 15: Length 5597:

Query Match	100.0%	Score 55371
Best Local Similarity	100.0%	Pred No. 0

Best Local Similarity	100.0%;	pred. No. 0;
Matches 5507.	Concervative	0.
	Mismatches	0.
	Indels	0.
	Gaps	0.

On 1 CGAGCCGAGCGGGCGTCGGTCCCGAGCTGCAATGAGCGGGCGCCCGGAGGGCTGTGACC 60

QY 1 CGGAGCCGAGCGGGCGTCCGTCCGCGGAGCTGCAATGAGCGGGCCCGGAGGCTGTGACC 60

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

QY 61 TGC GCG GCG GCG CCG ACC GGG CCC CTGA ATGG CGG CTG CTG AGG CGG CGG CGG C 120

[illegible]

121 GGGGGCTCAGGCTCTCGGGGGCGTGGCGTGGCGGTGAAGGGGTGATGCTGTTCAAGCTCC 180

Db 121 GCGGCTCAGGCTCCTCGGGCGTGGCGTGAGGGGTGATGCTGTTCAAGCTCC 180

Db	1361	CAATAGTAACACACACGAGATGTTTTTCGAAATTTTGAGCCACTTGCCCTACCTTTAGTTCCAC	1321
Qy	1321	CTGCTATTGAAAGTCACATTCATCGCATCGAAGGCGTGTCCAGAAGTGTATTATTACCTAA	1380
Db	1321	CTGCTATTGAAAGTCACATTCATCGCATCGAAGGCGTGTCCAGAAGTGTATTATTACCTAA	1380
Qy	1381	ATGATGATGTCACTGTTTGGGAAGGATGTCTGGCCAGATGATTTTTCAGTCACTCCAAAG	1440
Db	1381	ATGATGATGTCACTGTTTGGGAAGGATGTCTGGCCAGATGATTTTTCAGTCACTCCAAAG	1440
Qy	1441	GCCAGAAGGTTTATTTGACATGGCTGTGCCAAACTGTGCGAGGGGCTGCCCAGGTTCCCT	1500
Db	1441	GCCAGAAGGTTTATTTGACATGGCTGTGCCAAACTGTGCGAGGGGCTGCCCAGGTTCCCT	1500
Qy	1501	GGATTAAGGATGGCTATTGTGACAAGGCTGTGTAATTAATTCAGCTGCGATTTGGATGGTG	1560
Db	1501	GGATTAAGGATGGCTATTGTGACAAGGCTGTGTAATTAATTCAGCTGCGATTTGGATGGTG	1560
Qy	1561	GGGATTCCTCTGGAACACAGTGGAGGAGTCGCTATATTTCAGAGAGTGGAGTACTGGGA	1620
Db	1561	GGGATTCCTCTGGAACACAGTGGAGGAGTCGCTATATTTCAGAGAGTGGAGTACTGGGA	1620
Qy	1621	GTATTGGAGTTGGACACCCCTGGCAGCTTTCGTGGAGGAATAAACAGTGTCTCTTACTGTA	1680
Db	1621	GTATTGGAGTTGGACACCCCTGGCAGCTTTCGTGGAGGAATAAACAGTGTCTCTTACTGTA	1680
Qy	1681	ATCAGGAGTGGCAATTCCTGGCTGCTGATTAAGTCTGTGACCAAGCATGCAATGTCT	1740
Db	1681	ATCAGGAGTGGCAATTCCTGGCTGCTGATTAAGTCTGTGACCAAGCATGCAATGTCT	1740
Qy	1741	TGTCCTGTGGGTTTGATGCTGGCGACCTGTGGGCAAGATCAATTTTCATGAATGTATAAAG	1800
Db	1741	TGTCCTGTGGGTTTGATGCTGGCGACCTGTGGGCAAGATCAATTTTCATGAATGTATAAAG	1800
Qy	1801	TGATCCTTCTCCAAACACGACTCACTATATTATTCGAAAGTGAAATGCTGCTTATT	1860
Db	1801	TGATCCTTCTCCAAACACGACTCACTATATTATTCGAAAGTGAAATGCTGCTTATT	1860
Qy	1861	TCAGCTTTGCAGAAAGTAGCCAAAGAGGAGTCTGAAGTGCTCATAGTGACAAATCCCAATAA	1920
Db	1861	TCAGCTTTGCAGAAAGTAGCCAAAGAGGAGTCTGAAGTGCTCATAGTGACAAATCCCAATAA	1920
Qy	1921	TTGACATGTTCTTATTGCCAACAGATGGAAAAACCATCCACCTCATATAATGCAAGTGGAA	1980
Db	1921	TTGACATGTTCTTATTGCCAACAGATGGAAAAACCATCCACCTCATATAATGCAAGTGGAA	1980
Qy	1981	TGAATGCCACCAATACATTTTAATCTCAGTTTCAAATATCAAACGATGAGAGTTCA	2040
Db	1981	TGAATGCCACCAATACATTTTAATCTCAGTTTCAAATATCAAACGATGAGAGTTCA	2040
Qy	2041	AAATGTCAGATAACAGTGGAGGTGGACACAGGAGGAGCCAAACCTGAAATTCACGGCCC	2100
Db	2041	AAATGTCAGATAACAGTGGAGGTGGACACAGGAGGAGCCAAACCTGAAATTCACGGCCC	2100
Qy	2101	AGAAGGTTACGAAATTTAGTTAGTGTCCCATACCTCTTTCAGAGGGCGGAAATCGTTT	2160
Db	2101	AGAAGGTTACGAAATTTAGTTAGTGTCCCATACCTCTTTCAGAGGGCGGAAATCGTTT	2160
Qy	2161	TTGAGGATATCCCAAGAAAAACGCTTCGCCAAGTTTAAAGACATGATGTTAACTCAA	2220
Db	2161	TTGAGGATATCCCAAGAAAAACGCTTCGCCAAGTTTAAAGACATGATGTTAACTCAA	2220
Qy	2221	CAGAGGAGCCGAGGAGAGGTGAAATTTCCCTCTGTGTAATAATTCCTCTTCCAAAAG	2280
Db	2221	CAGAGGAGCCGAGGAGAGGTGAAATTTCCCTCTGTGTAATAATTCCTCTTCCAAAAG	2280
Qy	2281	ACGCCCATGTTGAGTCTCAATACCTTGAATTTGCAACTGGAAACATGAGACATCACCTTGA	2340
Db	2281	ACGCCCATGTTGAGTCTCAATACCTTGAATTTGCAACTGGAAACATGAGACATCACCTTGA	2340
Qy	2341	AAGGATACAAATTTGTCGAAGTCAGCCTTGCTGAGATCAATTTCTGATGAATCACAGATG	2400
Db	2341	AAGGATACAAATTTGTCGAAGTCAGCCTTGCTGAGATCAATTTCTGATGAATCACAGATG	2400







```
Db 4561 ATTGACATCTGAATGCTTTGTAAATTAATGACTTCAGCCCTTAAGAATGCTATGATTCA 4620
QY 4621 CGTGAGGCTTAATTTACAGGCTAGATTAGTACTACTACAGATGTAATTAATGTTT 4680
Db 4621 CGTGAGGCTTAATTTACAGGCTAGATTAGTACTACTACAGATGTAATTAATGTTT 4680
QY 4681 TGGAAATGTACATATTTCAACAGAGAGTGCTCATTTAGAAATGAGTAGTGCTGATGCA 4740
Db 4681 TGGAAATGTACATATTTCAACAGAGAGTGCTCATTTAGAAATGAGTAGTGCTGATGCA 4740
QY 4741 CTGGCACATTACAGTGGTGCTTTTAATACTCATTTGGTATATTTCCAGTAGTATCTCT 4800
Db 4741 CTGGCACATTACAGTGGTGCTTTTAATACTCATTTGGTATATTTCCAGTAGTATCTCT 4800
QY 4801 CTCAGTTGGTTTTTCATAGACAGAGGCGCAGCAACTTTCTTTGAAAAGGCTGTTAGT 4860
Db 4801 CTCAGTTGGTTTTTCATAGACAGAGGCGCAGCAACTTTCTTTGAAAAGGCTGTTAGT 4860
QY 4861 AAATTATTGCGAGGCACTGTGCTTTTGTATACATTTCTTCTGCTGTTTGTAGTTTGT 4920
Db 4861 AAATTATTGCGAGGCACTGTGCTTTTGTATACATTTCTTCTGCTGTTTGTAGTTTGT 4920
QY 4921 TTTTTCACAAACCCCTCTAAATATGTAATGTAATGTAATGTAATGTAATGTAATGTA 4980
Db 4921 TTTTTCACAAACCCCTCTAAATATGTAATGTAATGTAATGTAATGTAATGTAATGTA 4980
QY 4981 ACTGCCACACCCAGATGTGACCTCTGAGGCACTCATTTGCAATCACTGAGAAATATT 5040
Db 4981 ACTGCCACACCCAGATGTGACCTCTGAGGCACTCATTTGCAATCACTGAGAAATATT 5040
QY 5041 TTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 5100
Db 5041 TTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 5100
QY 5101 TGCAGTGGGCAATCTCAGCTCAGTCACTCGCCCTCCCGGTTCAAGCAGTTCTGTC 5160
Db 5101 TGCAGTGGGCAATCTCAGCTCAGTCACTCGCCCTCCCGGTTCAAGCAGTTCTGTC 5160
QY 5161 TCAGCTTCTGAGTGGGACTACAGTGCATGCACTCGCCCTCCCGGTTCAAGCAGTTCTG 5220
Db 5161 TCAGCTTCTGAGTGGGACTACAGTGCATGCACTCGCCCTCCCGGTTCAAGCAGTTCTG 5220
QY 5221 TTTTTCAGAGAGGAGGTTCCACCATATTTGGTCAAGCTTATCTTGAACCTCTGACCTC 5280
Db 5221 TTTTTCAGAGAGGAGGTTCCACCATATTTGGTCAAGCTTATCTTGAACCTCTGACCTC 5280
QY 5281 AGGTGATCCACCTGCTGCTCCCAAGTGTGAGATTACAGGCAATAGCCAGTGCAC 5340
Db 5281 AGGTGATCCACCTGCTGCTCCCAAGTGTGAGATTACAGGCAATAGCCAGTGCAC 5340
QY 5341 CCAGCGGAGATTAGTATTTTATGTTATGTTTAACTTGGCTGAGGCTAGCCATATTTATG 5400
Db 5341 CCAGCGGAGATTAGTATTTTATGTTATGTTTAACTTGGCTGAGGCTAGCCATATTTATG 5400
QY 5401 TCATAATCAATGAGTATTTGTAAGAGCAGATTCCATGAGTAACTCTGACAGGATTTTAG 5460
Db 5401 TCATAATCAATGAGTATTTGTAAGAGCAGATTCCATGAGTAACTCTGACAGGATTTTAG 5460
QY 5461 ATCATGATCTCAACAATATTTCTCCCAATGGCATACATCTTTTGTACAAAGAACTTGA 5520
Db 5461 ATCATGATCTCAACAATATTTCTCCCAATGGCATACATCTTTTGTACAAAGAACTTGA 5520
QY 5521 ATGTAATACGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 5580
Db 5521 ATGTAATACGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 5580
QY 5581 AAGTTAAATTTTGA 5597
Db 5581 AAGTTAAATTTTGA 5597
```

RESULT 6

```
US-10-024-197-3
; Sequence 3, Application US/10024197.
; Publication No. US20030133924A1
; GENERAL INFORMATION:
; APPLICANT: CAMFIELD, William
; TITLE OF INVENTION: HIGHLY PHOSPHORYLATED ACID BETA-GLUCOCEREBROSIDASE AND METHODS
; TITLE OF INVENTION: TREATING GAUCHER'S DISEASE
; FILE REFERENCE: 209794USO
; CURRENT APPLICATION NUMBER: US/10/024,197
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 5597
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-024-197-3
```

Query Match 100.0%; Score 5597; DB 15; Length 5597;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 5597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 CGGAGCGGCGGCGGCTCGTCCGCGAGCTGCAATGAGCGGCGCGGAGGCTGTGACC 60
Db 1 CGGAGCGGCGGCGGCTCGTCCGCGAGCTGCAATGAGCGGCGCGGAGGCTGTGACC 60
QY 61 TGGCGCGGCGGCGGCGGCGGCGGCGGCTGAAATGGCGGCTCGCTGAGCGGCGGCGGCGG 120
Db 61 TGGCGCGGCGGCGGCGGCGGCGGCGGCGGCTGAAATGGCGGCTCGCTGAGCGGCGGCGGCGG 120
QY 121 GCGGCGTCAAGCTCTCGGGGCGTGGGCTGGCGGTGAAGGGGTGATGCTGTTCAAGCTCC 180
Db 121 GCGGCGTCAAGCTCTCGGGGCGTGGGCTGGCGGTGAAGGGGTGATGCTGTTCAAGCTCC 180
QY 181 TGCAGAGACAACTATACCTGCTGCTCCACAGGTATGGGCTACGCTGTGCTTCTTGG 240
Db 181 TGCAGAGACAACTATACCTGCTGCTCCACAGGTATGGGCTACGCTGTGCTTCTTGG 240
QY 241 GCGTGTGTCACCATCGTCTCGCCCTTCAGTTCCGAGAGTGTTCTGGAATGAGCC 300
Db 241 GCGTGTGTCACCATCGTCTCGCCCTTCAGTTCCGAGAGTGTTCTGGAATGAGCC 300
QY 301 GAGATCAATACCATGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 360
Db 301 GAGATCAATACCATGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 360
QY 361 AGAATCGGCTTTGCTGCCCCATGCCGATTGACGTTGTTTACACCTGGGTGAATGGCACAG 420
Db 361 AGAATCGGCTTTGCTGCCCCATGCCGATTGACGTTGTTTACACCTGGGTGAATGGCACAG 420
QY 421 ATCTTGAACCTACGAGGAACTACAGCAGGTGAGAGACAGATGAGAGGAGGAGCAAG 480
Db 421 ATCTTGAACCTACGAGGAACTACAGCAGGTGAGAGACAGATGAGAGGAGGAGCAAG 480
QY 481 CAATGAGAGAAATCTTTGGGAAAAACAAACGGAACCTACTAAGAGAGGTGAGAAACAGT 540
Db 481 CAATGAGAGAAATCTTTGGGAAAAACAAACGGAACCTACTAAGAGAGGTGAGAAACAGT 540
QY 541 TAGAGTGTGTTGTAACACACTGCAATTAAGTGCGCAATGCTGTGACCCAGCCCTGCG 600
Db 541 TAGAGTGTGTTGTAACACACTGCAATTAAGTGCGCAATGCTGTGACCCAGCCCTGCG 600
QY 601 CAGCCAACTACCCCTGAAAGGAGTGCCTATCTTTTATCCTTCTTTTCAATTCGCCAGTG 660
Db 601 CAGCCAACTACCCCTGAAAGGAGTGCCTATCTTTTATCCTTCTTTTCAATTCGCCAGTG 660
QY 661 ACATTTTCAATGTTGCAAAAACCAAAAACCCCTTCTACCAATGCTCAGTTGTTGTTTGG 720
Db 661 ACATTTTCAATGTTGCAAAAACCAAAAACCCCTTCTACCAATGCTCAGTTGTTGTTTGG 720
QY 721 ACAGTACTAGGATGTTGAGATGCCACTCTGGACTGCTTAAAGAAATAGCAGACAGA 780
Db 721 ACAGTACTAGGATGTTGAGATGCCACTCTGGACTGCTTAAAGAAATAGCAGACAGA 780
```

QY 781 CAGTATGAGGGGTACTTGAACAACAGATGAAGAGTCCCTGGATTAGTCTAATGCAAG 840  
DB 781 CAGTATGAGGGGTACTTGAACAACAGATGAAGAGTCCCTGGATTAGTCTAATGCAAG 840  
QY 841 ATTTGGCTTTCTGAGTGGATTTCCACCAACATTCAGGAAACAAATCRAACTAARACAA 900  
DB 841 ATTTGGCTTTCTGAGTGGATTTCCACCAACATTCAGGAAACAAATCRAACTAARACAA 900  
QY 901 AATTGCCAGAAAAATCTTTCCTCTAAAGTCMAACTGTTGCAAGTGTATTCAGAGGCCAGTG 960  
DB 901 AATTGCCAGAAAAATCTTTCCTCTAAAGTCMAACTGTTGCAAGTGTATTCAGAGGCCAGTG 960  
QY 961 TAGCGCTTTCTAAACTGAAATACCCCAAGGATTTTCAAGAAATGGAATGAAGCAAACTAAGA 1020  
DB 961 TAGCGCTTTCTAAACTGAAATACCCCAAGGATTTTCAAGAAATGGAATGAAGCAAACTAAGA 1020  
QY 1021 AGAATCATGACCATTTGATGGAAGAACTGACCAATAGTCTGCAATATTTATTTGGGATC 1080  
DB 1021 AGAATCATGACCATTTGATGGAAGAACTGACCAATAGTCTGCAATATTTATTTGGGATC 1080  
QY 1081 TGAGCGCCATCAGCGAGTCTTAAGCAGGATGAAGACATCTCTGCGAGTCTGTTTGAAGATA 1140  
DB 1081 TGAGCGCCATCAGCGAGTCTTAAGCAGGATGAAGACATCTCTGCGAGTCTGTTTGAAGATA 1140  
QY 1141 ACGAGAACTGAGGTACTCATTTGCGATCTATCGAGAGCATGCAACATGGGTTGCGAATA 1200  
DB 1141 ACGAGAACTGAGGTACTCATTTGCGATCTATCGAGAGCATGCAACATGGGTTGCGAATA 1200  
QY 1201 TTTTCATTTGTCACCAACGGGAGATTCATCTCTGGCTGAACCTTGACAACTCTCGAGTGA 1260  
DB 1201 TTTTCATTTGTCACCAACGGGAGATTCATCTCTGGCTGAACCTTGACAACTCTCGAGTGA 1260  
QY 1261 CAATAGTAACACACAGAGATGTTTTCGAAATTTGAGCCACTTGCCTACCTTTAGTTTAC 1320  
DB 1261 CAATAGTAACACACAGAGATGTTTTCGAAATTTGAGCCACTTGCCTACCTTTAGTTTAC 1320  
QY 1321 CTGCTATTGAAAGTCACATTCATCGCATGAGGGCTGTCGCAAGTGTATTTTACCTAA 1380  
DB 1321 CTGCTATTGAAAGTCACATTCATCGCATGAGGGCTGTCGCAAGTGTATTTTACCTAA 1380  
QY 1381 ATGATGATGTCATGTTTGGGAAGATGTCGCGCAGATGATTTTACAGTCACTCCAAAG 1440  
DB 1381 ATGATGATGTCATGTTTGGGAAGATGTCGCGCAGATGATTTTACAGTCACTCCAAAG 1440  
QY 1441 GCCAGAGGTTTATTTGACATGGCTGTCGCAAACTGTCGCGAGGGCTGCCAGGTTCTCT 1500  
DB 1441 GCCAGAGGTTTATTTGACATGGCTGTCGCAAACTGTCGCGAGGGCTGCCAGGTTCTCT 1500  
QY 1501 GGATTAAGGATGGCTATTGTCAGAGGCTTGTATAATTCAGCTGCGATTTGGGATGGTG 1560  
DB 1501 GGATTAAGGATGGCTATTGTCAGAGGCTTGTATAATTCAGCTGCGATTTGGGATGGTG 1560  
QY 1561 GGGATGCTCTGGAACAGTGGAGGATGCTGCTATATTGCAAGGAGTGGAGTACTGGGA 1620  
DB 1561 GGGATGCTCTGGAACAGTGGAGGATGCTGCTATATTGCAAGGAGTGGAGTACTGGGA 1620  
QY 1621 GTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAAGTCTCTTACTGTA 1680  
DB 1621 GTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAAGTCTCTTACTGTA 1680  
QY 1681 ATCAGGATGTCGGAATTCCTGGCTGCTGATAGATTCGTGACCAAGCATGCAATGCTCT 1740  
DB 1681 ATCAGGATGTCGGAATTCCTGGCTGCTGATAGATTCGTGACCAAGCATGCAATGCTCT 1740  
QY 1741 TGTCTCTGGGTTTGTATGCTGGGACATGTCGCAAGATCAATTTTCATGAATTTGATAAAG 1800  
DB 1741 TGTCTCTGGGTTTGTATGCTGGGACATGTCGCAAGATCAATTTTCATGAATTTGATAAAG 1800  
QY 1801 TGATCCTTCTCCAAACACAGACTCACTATATTTTCCAAAGGATGCAATGCTGCTTATT 1860  
DB 1801 TGATCCTTCTCCAAACACAGACTCACTATATTTTCCAAAGGATGCAATGCTGCTTATT 1860

QY 1861 TCAGCTTTGCAGAGTAGCCAAAGAGGAGTTGAAGTGCCTATAGTGCACAAATCCAATAA 1920  
DB 1861 TCAGCTTTGCAGAGTAGCCAAAGAGGAGTTGAAGTGCCTATAGTGCACAAATCCAATAA 1920  
QY 1921 TTTGCAGATGCTTCTATTGCCAACAGTGGAAAAACCATCCACTCATAATGACAGTGGA 1980  
DB 1921 TTTGCAGATGCTTCTATTGCCAACAGTGGAAAAACCATCCACTCATAATGACAGTGGA 1980  
QY 1981 TGAATGCCACCAACAATATATTTTAATCTCAGTTTCAAAAATACAAAACATGAAGTTC 2040  
DB 1981 TGAATGCCACCAACAATATATTTTAATCTCAGTTTCAAAAATACAAAACATGAAGTTC 2040  
QY 2041 AAATGCAGATAACAGTGGAGTGGACACAAGGGAGGGACCAAACTGAATTTCTACGGGCC 2100  
DB 2041 AAATGCAGATAACAGTGGAGTGGACACAAGGGAGGGACCAAACTGAATTTCTACGGGCC 2100  
QY 2101 AGAAGGGTTACGAAAATTTAGTTAGTCCATAAACAATTTCCAGAGCGGAAATCCCTTT 2160  
DB 2101 AGAAGGGTTACGAAAATTTAGTTAGTCCATAAACAATTTCCAGAGCGGAAATCCCTTT 2160  
QY 2161 TTGAGGATATTTCCCAAGAAAACGCTTCCCGAGTTTAAAGAGACATGATTTAACTCAA 2220  
DB 2161 TTGAGGATATTTCCCAAGAAAACGCTTCCCGAGTTTAAAGAGACATGATTTAACTCAA 2220  
QY 2221 CAAGGAGAGCCAGGAAGAGTGAATAATTTCCCTGGTAAATATTTTCACTCTCTCCAAAG 2280  
DB 2221 CAAGGAGAGCCAGGAAGAGTGAATAATTTCCCTGGTAAATATTTTCACTCTCTCCAAAG 2280  
QY 2281 AGCCGAGTTGAGTCTCAATACCTTGGATTGGCAATCGAATCGAATCGAATCGAATCGA 2340  
DB 2281 AGCCGAGTTGAGTCTCAATACCTTGGATTGGCAATCGAATCGAATCGAATCGAATCGA 2340  
QY 2341 AAGGATACAAATTTGTCAGCTTGCAGATCATTTCTGATGAATCTCAGAGCATG 2400  
DB 2341 AAGGATACAAATTTGTCAGCTTGCAGATCATTTCTGATGAATCTCAGAGCATG 2400  
QY 2401 CTAAATAAAAATCAAGCTATAAACAAGTGAATAAACAAGTGAATAAACAAGTGAATAA 2460  
DB 2401 CTAAATAAAAATCAAGCTATAAACAAGTGAATAAACAAGTGAATAAACAAGTGAATAA 2460  
QY 2461 AGGAAACAGGTTTCATAAAGCACTTGGCAACAGCTTAGGAGTGTCTGAAGATTGC 2520  
DB 2461 AGGAAACAGGTTTCATAAAGCACTTGGCAACAGCTTAGGAGTGTCTGAAGATTGC 2520  
QY 2521 AGAGGTTGATCTTTCTGCAAGTGAATAAAGTGGTCAAGCAGGCTCAGAAATC 2580  
DB 2521 AGAGGTTGATCTTTCTGCAAGTGAATAAAGTGGTCAAGCAGGCTCAGAAATC 2580  
QY 2581 CACCCCTGAGCTTGGAGACCAAGCAAGATTTAGAGTGGAACTCACAACCAAAACCA 2640  
DB 2581 CACCCCTGAGCTTGGAGACCAAGCAAGATTTAGAGTGGAACTCACAACCAAAACCA 2640  
QY 2641 TAGGCGGAAATGTGCAAAAGAAAAAGCCCCATCTCTGATTGTTCCACTGGAAAGCCAGA 2700  
DB 2641 TAGGCGGAAATGTGCAAAAGAAAAAGCCCCATCTCTGATTGTTCCACTGGAAAGCCAGA 2700  
QY 2701 TGACAAAAGAAAAATCAGGGAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG 2760  
DB 2701 TGACAAAAGAAAAATCAGGGAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG 2760  
QY 2761 CTGAAAATCAGATAGCGGTTTACTGAAAGTGTACTTTGGAAGAAAGTGCAGCATTAACAG 2820  
DB 2761 CTGAAAATCAGATAGCGGTTTACTGAAAGTGTACTTTGGAAGAAAGTGCAGCATTAACAG 2820  
QY 2821 ATAGTTACTTGGGCTTTTGGCAATGGGAAAAAAAGTATTTTCCAGATCTTCTCGACG 2880  
DB 2821 ATAGTTACTTGGGCTTTTGGCAATGGGAAAAAAAGTATTTTCCAGATCTTCTCGACG 2880  
QY 2881 AAGAGAGTCAATTAAGACACAAATTTGGCATCTTCACTGATAGCAAAATTTCTGGAGGC 2940  
DB 2881 AAGAGAGTCAATTAAGACACAAATTTGGCATCTTCACTGATAGCAAAATTTCTGGAGGC 2940  
QY 2941 AACTAAAAGATACATTTGCAGATTTCCCTCAGATATGTAAATAAAATTTCTAAATAGCAAGT 3000

Db 2941 AACTAAAAGATACATTTGCGAGATCCCTCAGATATGTAATAATAATTTCTAATAATAGCAAGT 3000  
Qy 3001 TTGGATTACATCGCGGAAAGTCCCTGCTCAGTCCCTCAGATGATGAAACCGGATTTGTTA 3060  
Db 3001 TTGGATTACATCGCGGAAAGTCCCTGCTCAGTCCCTCAGATGATGAAACCGGATTTGTTA 3060  
Qy 3061 TGCAAGAACTCAAGATATGTTCCCTGAGAAATTTGCAAGACGTCATTTTCAACAAGTGC 3120  
Db 3061 TGCAAGAACTCAAGATATGTTCCCTGAGAAATTTGCAAGACGTCATTTTCAACAAGTGC 3120  
Qy 3121 GCCATTCTGAGGATATGAGATTTGCCCTTCTCTTATTTTATATCTCATGATGCACTGC 3180  
Db 3121 GCCATTCTGAGGATATGAGATTTGCCCTTCTCTTATTTTATATCTCATGATGCACTGC 3180  
Qy 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAGTTGATACAGATCAATCTGGTGTCTTGT 3240  
Db 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAGTTGATACAGATCAATCTGGTGTCTTGT 3240  
Qy 3241 CTGACAGAAATCCGAACACTGGCTACCGAAATTCACGAACTGCCGTTAAGTTTGCAGG 3300  
Db 3241 CTGACAGAAATCCGAACACTGGCTACCGAAATTCACGAACTGCCGTTAAGTTTGCAGG 3300  
Qy 3301 ATTTGACAGTCTGGAACACATGCTATATAAATTTGCTCAAAAAATGCTTCTCTGCTGATATCA 3360  
Db 3301 ATTTGACAGTCTGGAACACATGCTATATAAATTTGCTCAAAAAATGCTTCTCTGCTGATATCA 3360  
Qy 3361 CGCAGCTAAATAATTTCCACCACTCAGGAATCCTACTATGATCCCAACCTGCCACCGG 3420  
Db 3361 CGCAGCTAAATAATTTCCACCACTCAGGAATCCTACTATGATCCCAACCTGCCACCGG 3420  
Qy 3421 TCACATAAAGTCTAGTAACAACTGTAAACGAGTAATCTGACAAATCCCAAGGCAATATA 3480  
Db 3421 TCACATAAAGTCTAGTAACAACTGTAAACGAGTAATCTGACAAATCCCAAGGCAATATA 3480  
Qy 3481 AGGACAAAAACAAATATAGTTTGAATCATGGSAGAGAGAAATCGCTTTTAAATGA 3540  
Db 3481 AGGACAAAAACAAATATAGTTTGAATCATGGSAGAGAGAAATCGCTTTTAAATGA 3540  
Qy 3541 TTGCTACCAAGTTTCTCATGTGTTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 3600  
Db 3541 TTGCTACCAAGTTTCTCATGTGTTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 3600  
Qy 3601 AGTTTGTTCCTGGAATGACAACTTGCACCAATCATAAAGATGCTCAGACAGTGAAGG 3660  
Db 3601 AGTTTGTTCCTGGAATGACAACTTGCACCAATCATAAAGATGCTCAGACAGTGAAGG 3660  
Qy 3661 CTGTTCTCAGGACTTCTATGAATCCATGTTCCCAATACCTTCCCAATTTGAATGCA 3720  
Db 3661 CTGTTCTCAGGACTTCTATGAATCCATGTTCCCAATACCTTCCCAATTTGAATGCA 3720  
Qy 3721 GAGAGTATCGAAACCGTTTCTCATGATGATGATGATGATGATGATGATGATGATGATGAT 3780  
Db 3721 GAGAGTATCGAAACCGTTTCTCATGATGATGATGATGATGATGATGATGATGATGATGAT 3780  
Qy 3781 ACAAAATGAAGTTTGGACCATTTGCTAGTATGATGATGATGATGATGATGATGATGATGAT 3840  
Db 3781 ACAAAATGAAGTTTGGACCATTTGCTAGTATGATGATGATGATGATGATGATGATGATGAT 3840  
Qy 3841 CATTTTTCGAGCAGTTAATGCACTTAAGCGGAGATATTTCCAGAGAGAGGATAC 3900  
Db 3841 CATTTTTCGAGCAGTTAATGCACTTAAGCGGAGATATTTCCAGAGAGAGGATAC 3900  
Qy 3901 ACAAGAGCTAGTCCCAATCGATCAGATATGATGATGATGATGATGATGATGATGATGATGAT 3960  
Db 3901 ACAAGAGCTAGTCCCAATCGATCAGATATGATGATGATGATGATGATGATGATGATGATGAT 3960  
Qy 3961 CCTCAGCATTTACTGAGCATTTTAAACTCAGCTTCCAGAGATGCTTTTGTGATGATGAT 4020  
Db 3961 CCTCAGCATTTACTGAGCATTTTAAACTCAGCTTCCAGAGATGCTTTTGTGATGATGAT 4020  
Qy 4021 GCTTAGCAGTTTGGCCGAGAGAGAAATATCCAGTACCATGCTGTTTGTGGCATGAA 4080

Db 4021 GCTTAGCAGTTTGGCCGAGAGAGAAATATATCCAGTACCATGCTGCTGTTTGTGGCATGAA 4080  
Qy 4081 TATAGCCCACTGACTAGGAATTTATTAACCAACCCACTGAAACCTTGTGTGCGAGCAGC 4140  
Db 4081 TATAGCCCACTGACTAGGAATTTATTAACCAACCCACTGAAACCTTGTGTGCGAGCAGC 4140  
Qy 4141 TCTGAACTCATTTTACTTTTAAAGAAATTTGCTCATGGACCTGCTCATCTTTTATATAAAA 4200  
Db 4141 TCTGAACTCATTTTACTTTTAAAGAAATTTGCTCATGGACCTGCTCATCTTTTATATAAAA 4200  
Qy 4201 GGCTCATCTGACAGAGACAGCTGTTTAAATTTCCACAGCAATCATTTGCAGACTAACTTTAT 4260  
Db 4201 GGCTCATCTGACAGAGACAGCTGTTTAAATTTCCACAGCAATCATTTGCAGACTAACTTTAT 4260  
Qy 4261 TAGGAGAACTTATGCCAGCTGGGAGTATGCTTAAGAGGCTCCAGTCTTTCCATTTCCAA 4320  
Db 4261 TAGGAGAACTTATGCCAGCTGGGAGTATGCTTAAGAGGCTCCAGTCTTTCCATTTCCAA 4320  
Qy 4321 AGCCTTTTCTCTAAAGTTTTCGCACTTTTTCATTTTCCATTTTCCATTTTAAAGTAGTACTA 4380  
Db 4321 AGCCTTTTCTCTAAAGTTTTCGCACTTTTTCATTTTCCATTTTCCATTTTAAAGTAGTACTA 4380  
Qy 4381 AGTTAACTAGTTATTTCTGCTTCTGAGTATAACGAATTTGGGATGCTTAAACCTATTTTAA 4440  
Db 4381 AGTTAACTAGTTATTTCTGCTTCTGAGTATAACGAATTTGGGATGCTTAAACCTATTTTAA 4440  
Qy 4441 TAGATGTTATTTTAAATAATGACAGCAATATCACCTCTTATTGACAATACCTAAATATGAG 4500  
Db 4441 TAGATGTTATTTTAAATAATGACAGCAATATCACCTCTTATTGACAATACCTAAATATGAG 4500  
Qy 4501 TTTTATTAATATTTAAGACTGTPAAATGGTCTTAAACCACTAACTACTGAAGAGCTCAATG 4560  
Db 4501 TTTTATTAATATTTAAGACTGTPAAATGGTCTTAAACCACTAACTACTGAAGAGCTCAATG 4560  
Qy 4561 ATTGACATCTGAAATGCTTTGTAATTTGACTTCCAGCCCTTAAGAAATGCTATGATTTCA 4620  
Db 4561 ATTGACATCTGAAATGCTTTGTAATTTGACTTCCAGCCCTTAAGAAATGCTATGATTTCA 4620  
Qy 4621 CGTGAGGCTCTAATTTCAACAGGCTAGAGTTAGTACTACTTACCAGATGTAATATGTTT 4680  
Db 4621 CGTGAGGCTCTAATTTCAACAGGCTAGAGTTAGTACTACTTACCAGATGTAATATGTTT 4680  
Qy 4681 TGAAATGTACATATTTCAACAGAGTGCCTCATTTTAGAAATGAGTAGTCTGATGGCA 4740  
Db 4681 TGAAATGTACATATTTCAACAGAGTGCCTCATTTTAGAAATGAGTAGTCTGATGGCA 4740  
Qy 4741 CTGGCACAATACAGTGGTGTCTTTTAACTACTCTGTTATTTTCCAGTAGTACTCTCT 4800  
Db 4741 CTGGCACAATACAGTGGTGTCTTTTAACTACTCTGTTATTTTCCAGTAGTACTCTCT 4800  
Qy 4801 CTCAGTTGGTTTGTATGAGAACAGAGCCAGCAAACTTTCTTTGTAAAGGCTGGTTAGT 4860  
Db 4801 CTCAGTTGGTTTGTATGAGAACAGAGCCAGCAAACTTTCTTTGTAAAGGCTGGTTAGT 4860  
Qy 4861 AAATTTATTCAGGCGCACCTGCTGTTGTCATACATCTCTTCTGCTGTTTGTAGTTGT 4920  
Db 4861 AAATTTATTCAGGCGCACCTGCTGTTTGTGTCATACATCTCTTCTGCTGTTTGTAGTTGT 4920  
Qy 4921 TTTTTCCTCAAAACCCCTCTAAATAATGTAATAAATCCATGTTTGTAGCTTGCAGCTGTACAAA 4980  
Db 4921 TTTTTCCTCAAAACCCCTCTAAATAATGTAATAAATCCATGTTTGTAGCTTGCAGCTGTACAAA 4980  
Qy 4981 ACTGCCACCCAGCCAGATGTCACCTCAGGCGCATCTTTGCCAATCACTGAGAAATTTT 5040  
Db 4981 ACTGCCACCCAGCCAGATGTCACCTCAGGCGCATCTTTGCCAATCACTGAGAAATTTT 5040  
Qy 5041 TTGTTTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 5100  
Db 5041 TTGTTTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 5100  
Qy 5101 TGCACTGGCGCAATCTCAGCTCAGTCACTGCAACCTCCGCTCCCGGTTCAAGCAGTTCTGTC 5160  
Db 5101 TGCACTGGCGCAATCTCAGCTCAGTCACTGCAACCTCCGCTCCCGGTTCAAGCAGTTCTGTC 5160



QY 1321 CTGCTATTGAAAGTCACATTCATCGCATCGAAGGCTGTCCAGAACTTTATTACCTAA 1380  
DB 1321 CTGCTATTGAAAGTCACATTCATCGCATCGAAGGCTGTCCAGAACTTTATTACCTAA 1380  
QY 1381 ATGATGATGATCATCTTTGGGAAGATGCTCTGGCCAGATGATTTTACAGTCACCTCCAAAG 1440  
DB 1381 ATGATGATGATCATCTTTGGGAAGATGCTCTGGCCAGATGATTTTACAGTCACCTCCAAAG 1440  
QY 1441 GCCAAGAGGTTTATTGACATGGGCTGTGCCAAACTGTGCCAGGGGTGCCAGGTTCCCT 1500  
DB 1441 GCCAAGAGGTTTATTGACATGGGCTGTGCCAAACTGTGCCAGGGGTGCCAGGTTCCCT 1500  
QY 1501 GGATTAAGGATGCTATTGTGACAAAGCTTTGTAATAATTACAGCTCGGATGGATGGTG 1560  
DB 1501 GGATTAAGGATGCTATTGTGACAAAGCTTTGTAATAATTACAGCTCGGATGGATGGTG 1560  
QY 1561 GGGATTGCTCTGGAAAACAGTGGAGGAGTGCCTATATTGACAGAGTGGAGTACTGGGA 1620  
DB 1561 GGGATTGCTCTGGAAAACAGTGGAGGAGTGCCTATATTGACAGAGTGGAGTACTGGGA 1620  
QY 1621 GTATTGGAGTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAAGTCTCTTACTGTA 1680  
DB 1621 GTATTGGAGTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAAGTCTCTTACTGTA 1680  
QY 1681 ATCAGGATGTGGAAATTCCTGGCTCGCTGATAGTTCTGTGACCAAGCATGCAATGTCT 1740  
DB 1681 ATCAGGATGTGGAAATTCCTGGCTCGCTGATAGTTCTGTGACCAAGCATGCAATGTCT 1740  
QY 1741 TGTCTGTGGTTTGTAGTCTGGGACCTGTGGGCAAGATCAATTTTCATGAATTTGATAAAG 1800  
DB 1741 TGTCTGTGGTTTGTAGTCTGGGACCTGTGGGCAAGATCAATTTTCATGAATTTGATAAAG 1800  
QY 1801 TGATCTCTTCCMAACAGACTCACTATATTATTCMAAAGTGAATGCTGCTCTATT 1860  
DB 1801 TGATCTCTTCCMAACAGACTCACTATATTATTCMAAAGTGAATGCTGCTCTATT 1860  
QY 1861 TCAGCTTTGAGAAAGTACCAAGAGGAGTGTGAAGTGTCTATAGTGAACAATCCAAATA 1920  
DB 1861 TCAGCTTTGAGAAAGTACCAAGAGGAGTGTGAAGTGTCTATAGTGAACAATCCAAATA 1920  
QY 1921 TTGACATGCTTCTATTGGCCAAACAGTGGAAAACCTTCACTCATATGCAAGTGGAA 1980  
DB 1921 TTGACATGCTTCTATTGGCCAAACAGTGGAAAACCTTCACTCATATGCAAGTGGAA 1980  
QY 1981 TGAATGCCACCAACAATACATTTTAATCTCAGTTTCAAAATACAAACGATGAAGATTCA 2040  
DB 1981 TGAATGCCACCAACAATACATTTTAATCTCAGTTTCAAAATACAAACGATGAAGATTCA 2040  
QY 2041 AAATGCAGATAACAGTGGAGTGGACACAAGGAGGAGCAAAACTCAATTTACCGCCC 2100  
DB 2041 AAATGCAGATAACAGTGGAGTGGACACAAGGAGGAGCAAAACTCAATTTACCGCCC 2100  
QY 2101 AGAAGGTTACGAAAATTTAGTTAGTCCCATACACTTCTTCCAGAGGGGGAATCCTTT 2160  
DB 2101 AGAAGGTTACGAAAATTTAGTTAGTCCCATACACTTCTTCCAGAGGGGGAATCCTTT 2160  
QY 2161 TTGAGGATATTCCAAAGAAAACCGCTTCCGAAAGTTTAAGACATGATTTAACTCAA 2220  
DB 2161 TTGAGGATATTCCAAAGAAAACCGCTTCCGAAAGTTTAAGACATGATTTAACTCAA 2220  
QY 2221 CAAGGAGAGCCAGGAAGAGTGAATAATTCCTCGTGAATAATTTCATCTCTTCCAAAG 2280  
DB 2221 CAAGGAGAGCCAGGAAGAGTGAATAATTCCTCGTGAATAATTTCATCTCTTCCAAAG 2280  
QY 2281 ACGCCAGTTGAGTCTCAATACCTTGGATTGCAACTTGAACATGGAGACATCACTTTGA 2340  
DB 2281 ACGCCAGTTGAGTCTCAATACCTTGGATTGCAACTTGAACATGGAGACATCACTTTGA 2340  
QY 2341 AAGGATACAATTTGTCCAAGTCAGCTTGTGAGATCAATTTCTGATGAATCAAGCATG 2400  
DB 2341 AAGGATACAATTTGTCCAAGTCAGCTTGTGAGATCAATTTCTGATGAATCAAGCATG 2400  
QY 2401 CTAATAATAAATAACAGCTTATAATAACATGAAACAATGACAGTTTGGTGGCTCCAC 2460

DB 2401 CTAATAATAAATAACAGCTTATAATAACATGAAACAATGACAGTTTGGTGGCTCCAC 2460  
QY 2461 AGSAAAACAGGTTTCATTAAGCATCTTCCAAAACAGCTTAGGAGTCTGGAAGATTGC 2520  
DB 2461 AGSAAAACAGGTTTCATTAAGCATCTTCCAAAACAGCTTAGGAGTCTGGAAGATTGC 2520  
QY 2521 AGAGGTTGACTTTTCTCTGAGTGTAAAGTGAATGGTCAATGACAGGTCAGATC 2580  
DB 2521 AGAGGTTGACTTTTCTCTGAGTGTAAAGTGAATGGTCAATGACAGGTCAGATC 2580  
QY 2581 CACCCCTGGACTTTGAGAGCCACAGCAAGATTTAGAGTGAAGAACTCACACCCAAAACCA 2640  
DB 2581 CACCCCTGGACTTTGAGAGCCACAGCAAGATTTAGAGTGAAGAACTCACACCCAAAACCA 2640  
QY 2641 TAGGCGGAAATGTGACAAAAGAAAAGCCCATCTCTGATTTGTTCTCACTGGAAGCCAGA 2700  
DB 2641 TAGGCGGAAATGTGACAAAAGAAAAGCCCATCTCTGATTTGTTCTCACTGGAAGCCAGA 2700  
QY 2701 TGACAAAAGAAAAGAAATCAAGGGAAGAAAAGAGAACAGTAGTAGAATGGAGGAAAATG 2760  
DB 2701 TGACAAAAGAAAAGAAATCAAGGGAAGAAAAGAGAACAGTAGTAGAATGGAGGAAAATG 2760  
QY 2761 CTGAAAATCACATAGCGTTTACTGAAGTGTACTTTGGAAGAAAGCTGCAGCATTTACACAG 2820  
DB 2761 CTGAAAATCACATAGCGTTTACTGAAGTGTACTTTGGAAGAAAGCTGCAGCATTTACACAG 2820  
QY 2821 ATAGTTTACTTGGGCTTTTGGCCATGGGAGAAAAGATTTTCCAAGATCTTCTCGAGC 2880  
DB 2821 ATAGTTTACTTGGGCTTTTGGCCATGGGAGAAAAGATTTTCCAAGATCTTCTCGAGC 2880  
QY 2881 AGAAGAGTCAATTTGAAGACACAATTTGGCATACTTCACTGATAGCAAAAATTTCTGGAGGC 2940  
DB 2881 AGAAGAGTCAATTTGAAGACACAATTTGGCATACTTCACTGATAGCAAAAATTTCTGGAGGC 2940  
QY 2941 AACTAAAAGATACATTTGCAGATTTCCCTCAGATATGTAATAAATAATTTCTAAATAGCAAGT 3000  
DB 2941 AACTAAAAGATACATTTGCAGATTTCCCTCAGATATGTAATAAATAATTTCTAAATAGCAAGT 3000  
QY 3001 TTGGATTACATCGCGGAAAGTCCCTGCTCAGTCCTCACATGATGTAATGTAATTTGACCGGATTGTA 3060  
DB 3001 TTGGATTACATCGCGGAAAGTCCCTGCTCAGTCCTCACATGATGTAATGTAATTTGTA 3060  
QY 3061 TGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGCAAGACGTCATTTCAAAAAGTGC 3120  
DB 3061 TGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGCAAGACGTCATTTCAAAAAGTGC 3120  
QY 3121 GCCATCTGAGGATATGAGTTTGCCTTCTCTTATTTTATTTATCTCATAGTGCAGTGC 3180  
DB 3121 GCCATCTGAGGATATGAGTTTGCCTTCTCTTATTTTATTTATCTCATAGTGCAGTGC 3180  
QY 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTCTTGT 3240  
DB 3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTCTTGT 3240  
QY 3241 CTGACAGAGAAATCCGAACACTGGCTACAGAAATTCAGAACTGCCGTTAAGTTTGCAGG 3300  
DB 3241 CTGACAGAGAAATCCGAACACTGGCTACAGAAATTCAGAACTGCCGTTAAGTTTGCAGG 3300  
QY 3301 ATTTGACAGGTTCTGGAACACATGCTAATAATTGCTCAAAAATGCTTCTGCTGATATCA 3360  
DB 3301 ATTTGACAGGTTCTGGAACACATGCTAATAATTGCTCAAAAATGCTTCTGCTGATATCA 3360  
QY 3361 GCGAGTAAATAATTTCCAACTCAGGAAATCTCTACTATGATCCCAACCTGCCACCGG 3420  
DB 3361 GCGAGTAAATAATTTCCAACTCAGGAAATCTCTACTATGATCCCAACCTGCCACCGG 3420  
QY 3421 TCATTAAGTCTAGTAAACAACTGTAAAACAGTAACTGACAAAATCCAAAAGCATATA 3480  
DB 3421 TCATTAAGTCTAGTAAACAACTGTAAAACAGTAACTGACAAAATCCAAAAGCATATA 3480  
QY 3481 AGGACAAAACAAATATAGGTTTGAATTCATGGGAGAGAAATGCTTTTAAAATGA 3540



3481	AGGACAAAA	CAATATATAGGTTTGAATCATGGGAGAGMAGAAATCGCTTTTAAATATGA	3544
3541	TTCGTACCAC	GTGTTTCTCATGTGGTTGGCCAGTTGGATGACATAAAGAAAAAACCCCTAGGA	3600
3541	TTCGTACCAC	AGTTTCTCATGTGGTTGGCCAGTTGGATGACATAAAGAAAAAACCCCTAGGA	3600
3601	AGTTTGT	TTCCCTGGAATGACAACTTGA	3660
3601	AGTTTGT	TTCCCTGGAATGACAACTTGA	3660
3661	CTGTTCTCAGG	AGCTTCTATGAATCCATGTTCCCCATACCTTCCCAATTTGAACCTGCCAA	3720
3661	CTGTTCTCAGG	AGCTTCTATGAATCCATGTTCCCCATACCTTCCCAATTTGAACCTGCCAA	3720
3721	GAGAGTATCGAA	ACCGTTCCCTTCATATGATGAGCTGCAGGAATGAGGGCTTATTCGAG	3780
3721	GAGAGTATCGAA	ACCGTTCCCTTCATATGATGAGCTGCAGGAATGAGGGCTTATTCGAG	3780
3781	ACAAATGAAG	TTTGGACCCATTGTTACTAGCACATTTGATTTGTTTACTATATCT	3840
3781	ACAAATGAAG	TTTGGACCCATTGTTACTAGCACATTTGATTTGTTTACTATATCT	3840
3841	CAITTTTTTGCTGAG	CAGCTTAATTCGACTTAAGCGGAGATTTTCCAGAAAGGAGATAC	3900
3841	CAITTTTTTGCTGAG	CAGCTTAATTCGACTTAAGCGGAGATTTTCCAGAAAGGAGATAC	3900
3901	ACAAAGAGCTAGT	CCCAATCGAATCAGAGTATAGAATCTTCATTTGAACCACTCTA	3960
3901	ACAAAGAGCTAGT	CCCAATCGAATCAGAGTATAGAATCTTCATTTGAACCACTCTA	3960
3961	CCTCAGCATTTA	CTGAGCATTTTAAACTCAGCTTACAGAGATGCTTTTGTGATGTGAT	4020
3961	CCTCAGCATTTA	CTGAGCATTTTAAACTCAGCTTACAGAGATGCTTTTGTGATGTGAT	4020
4021	GCITTAGCAGTTT	GGCCGAGAAAGAAATATACGATACATGCTGTTTGTGGCATGAA	4080
4021	GCITTAGCAGTTT	GGCCGAGAAAGAAATATACGATACATGCTGTTTGTGGCATGAA	4080
4081	TATAGCCCACTGACTAG	GAATTTTAAACCAACCCACTGAAAACTTGTGTGCTCGAGCAGC	4140
4081	TATAGCCCACTGACTAG	GAATTTTAAACCAACCCACTGAAAACTTGTGTGCTCGAGCAGC	4140
4141	TCTGAAC	CTGATTTTACTPTTTTAAAGAAATTTGCTCATGGA	4200
4141	TCTGAAC	CTGATTTTACTPTTTTAAAGAAATTTGCTCATGGA	4200
4201	GGCTCACTGACAGAGAC	AGCTGTTAATTTCCACAGCAATCATGCGACTCACTTAT	4260
4201	GGCTCACTGACAGAGAC	AGCTGTTAATTTCCACAGCAATCATGCGACTCACTTAT	4260
4261	TAGGAGAAGCCTAT	GCAGCTGGGAGTGATTTGCTTAAGAGGCTCCAGTCTTTGCACTCCAA	4320
4261	TAGGAGAAGCCTAT	GCAGCTGGGAGTGATTTGCTTAAGAGGCTCCAGTCTTTGCACTCCAA	4320
4321	AGCCTTTTGCTAAAG	TTTTCGACTTTTTCATTTCCCATTTTAAAGTAGTACTTA	4380
4321	AGCCTTTTGCTAAAG	TTTTCGACTTTTTCATTTCCCATTTTAAAGTAGTACTTA	4380
4381	AGTTAACTAGTAT	TTCTTGCTTGAATTAACGAATTTGGATGCTTAAACCTATTTTA	4440
4381	AGTTAACTAGTAT	TTCTTGCTTGAATTAACGAATTTGGATGCTTAAACCTATTTTA	4440
4441	TAGATGTTATTTAAAT	TAATGACCAATATCACCTTTTGA	4500
4441	TAGATGTTATTTAAAT	TAATGACCAATATCACCTTTTGA	4500
4501	TTTATTAATTTTAAAG	CTGAAATGGCTTTTAAACCACTTAACTGAGAGCTCAATG	4560
4501	TTTATTAATTTTAAAG	CTGAAATGGCTTTTAAACCACTTAACTGAGAGCTCAATG	4560
4561	ATTGACATCTGAAAT	TGCTTTGTAATTTATGACTTTCAGCCCTTAAGAAATGCTATGATTTCA	4620
4561	ATTGACATCTGAAAT	TGCTTTGTAATTTATGACTTTCAGCCCTTAAGAAATGCTATGATTTCA	4620





1861 TCAGCTTTGAGAGTAGCCCAAGAGGAGTTGAAGGTGCCTATAGTGACAAATCCAAATA 1920 QY  
1861 TCAGCTTTGAGAGTAGCCCAAGAGGAGTTGAAGGTGCCTATAGTGACAAATCCAAATA 1920 Db  
1921 TTGACATGCTTCTATTGCGCAACAGTAGGAAACCAATCCACCTCATPAATGCAAGTGGAA 1980 QY  
1921 TTGACATGCTTCTATTGCGCAACAGTAGGAAACCAATCCACCTCATPAATGCAAGTGGAA 1980 Db  
1981 TGAATGCCACCAATACATTTTAACTCTACGTTTCAAAATCAAAACGATGAAGTTC 2040 QY  
1981 TGAATGCCACCAATACATTTTAACTCTACGTTTCAAAATCAAAACGATGAAGTTC 2040 Db  
2041 AAATGACAGATAACAGTGGAGTGGACACAAAGGAGGAGCAAAACCTGAATTTCTACGGCCC 2100 QY  
2041 AAATGACAGATAACAGTGGAGTGGACACAAAGGAGGAGCAAAACCTGAATTTCTACGGCCC 2100 Db  
2101 AGAAGGTTACGAAATTTAGTTAGTCCCATTAACACTTCTTCAGAGGCGGAAATCCCTTT 2160 QY  
2101 AGAAGGTTACGAAATTTAGTTAGTCCCATTAACACTTCTTCAGAGGCGGAAATCCCTTT 2160 Db  
2161 TTGAGGATATTCCCAAGAAAACGCTTCCCGAAGTTTAAAGACATGATGTTAACTCAA 2220 QY  
2161 TTGAGGATATTCCCAAGAAAACGCTTCCCGAAGTTTAAAGACATGATGTTAACTCAA 2220 Db  
2221 CAAGGAGAGCCAGGAGAGGTGAAATTTCCCTGTTAAATATTTCACTCTCTTCCAAAAG 2280 QY  
2221 CAAGGAGAGCCAGGAGAGGTGAAATTTCCCTGTTAAATATTTCACTCTCTTCCAAAAG 2280 Db  
2281 ACGCCAGTTGAGTCTCAATACCTTGGATTGCAACTGGAACTGAGACATGACATCTTGA 2340 QY  
2281 ACGCCAGTTGAGTCTCAATACCTTGGATTGCAACTGGAACTGAGACATGACATCTTGA 2340 Db  
2341 AAGGATACAAATTTGTCAGCTGAGCTTGTGAGATCATTTCTGATGAATCTCACAGCATG 2400 QY  
2341 AAGGATACAAATTTGTCAGCTGAGCTTGTGAGATCATTTCTGATGAATCTCACAGCATG 2400 Db  
2401 CTAATAATAAATCAAGCTATTAATACAGATGAACAAATGACAGTTTGGTGGCTCCAC 2460 QY  
2401 CTAATAATAAATCAAGCTATTAATACAGATGAACAAATGACAGTTTGGTGGCTCCAC 2460 Db  
2461 AGGAAAACAGGTTCTATAAAGCATCTTGCCAAACAGCTTAGAGTGTCTGAAAAGATTGC 2520 QY  
2461 AGGAAAACAGGTTCTATAAAGCATCTTGCCAAACAGCTTAGAGTGTCTGAAAAGATTGC 2520 Db  
2521 AGAGTTGACTTTTCTGCGAGTGAATGATGATGATGATGATGATGATGATGATGATGATG 2580 QY  
2521 AGAGTTGACTTTTCTGCGAGTGAATGATGATGATGATGATGATGATGATGATGATGATG 2580 Db  
2581 CACCCCTGGAATTTGGAGACACAGCAAGATTAGAGTGGAACTCACACCCCAAAAACCA 2640 QY  
2581 CACCCCTGGAATTTGGAGACACAGCAAGATTAGAGTGGAACTCACACCCCAAAAACCA 2640 Db  
2641 TAGCCGAAATGTGACAAAAGAAAAGCCCAATCTCTGATGTTGTTCCACTGGAAGCCAGA 2700 QY  
2641 TAGCCGAAATGTGACAAAAGAAAAGCCCAATCTCTGATGTTGTTCCACTGGAAGCCAGA 2700 Db  
2701 TGACAAAGAAAAGAAAATCACAGGGAAGAAAAGAGAACAGACAGTGAATGGAGAAAATG 2760 QY  
2701 TGACAAAGAAAAGAAAATCACAGGGAAGAAAAGAGAACAGACAGTGAATGGAGAAAATG 2760 Db  
2761 CTGAAAATCATAGGCGTTTACTGAAAGTGTACTTGGAAAGAAAGCTGACAGCATTTACACAG 2820 QY  
2761 CTGAAAATCATAGGCGTTTACTGAAAGTGTACTTGGAAAGAAAGCTGACAGCATTTACACAG 2820 Db  
2821 ATAGTTACTTTGGGCTTTTCCCATGGAGAAAAGAAATTTTCCAAAGATCTTCTCGAGC 2880 QY  
2821 ATAGTTACTTTGGGCTTTTCCCATGGAGAAAAGAAATTTTCCAAAGATCTTCTCGAGC 2880 Db  
2881 AAGAGAGTCAATGAGACAAATTTGGCATACTTCACTGATAGCAAAAATACTGGGAGGC 2940 QY  
2881 AAGAGAGTCAATGAGACAAATTTGGCATACTTCACTGATAGCAAAAATACTGGGAGGC 2940 Db

2941 AACTAAAGATACATTTGCGAGATTCCTCAGATATGTAAATAAAATTTCTAAATAGCAAGT 3000 QY  
2941 AACTAAAGATACATTTGCGAGATTCCTCAGATATGTAAATAAAATTTCTAAATAGCAAGT 3000 Db  
3001 TTGATTTCAATTCGCGAAAGTCCCTGCTCAGTCCCTCAGTCCCTCAGTCCCTGATGACCGGATTTGTTA 3060 QY  
3001 TTGATTTCAATTCGCGAAAGTCCCTGCTCAGTCCCTCAGTCCCTCAGTCCCTGATGACCGGATTTGTTA 3060 Db  
3061 TGCAAGAACTGCAAGATATGTTCCCTGAAGATTTGCAAGAGAGTCAATTTCAAAAGTGC 3120 QY  
3061 TGCAAGAACTGCAAGATATGTTCCCTGAAGATTTGCAAGAGAGTCAATTTCAAAAGTGC 3120 Db  
3121 GCCATTTCTGAGGATATGAGTGGTTCCTTCTTATTTTATTTATCTCATGAGTGCAGTGC 3180 QY  
3121 GCCATTTCTGAGGATATGAGTGGTTCCTTCTTATTTTATTTATCTCATGAGTGCAGTGC 3180 Db  
3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTGTCTTGT 3240 QY  
3181 AGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTGTCTTGT 3240 Db  
3241 CTGACAGAGAAATCCGAACACTGGCTACCAAGATTTCAAGAACTGCCGTTAAAGTTGCGAG 3300 QY  
3241 CTGACAGAGAAATCCGAACACTGGCTACCAAGATTTCAAGAACTGCCGTTAAAGTTGCGAG 3300 Db  
3301 ATTTGACAGGCTGGAAACACATGCTAATTAATTTGCTCAAAATGCTTCCCTGCTGATATCA 3360 QY  
3301 ATTTGACAGGCTGGAAACACATGCTAATTAATTTGCTCAAAATGCTTCCCTGCTGATATCA 3360 Db  
3361 CGCAGCTAAATTAATTTCCACCAACTCAGGAATCTTACTATGATCCCAACCTGCCACCGG 3420 QY  
3361 CGCAGCTAAATTAATTTCCACCAACTCAGGAATCTTACTATGATCCCAACCTGCCACCGG 3420 Db  
3421 TCACTAAAAGTCTAGTAACAACTGTAAACAGTAAACAGTAAACAGTAAACAGTAAACAGTAA 3480 QY  
3421 TCACTAAAAGTCTAGTAACAACTGTAAACAGTAAACAGTAAACAGTAAACAGTAAACAGTAA 3480 Db  
3481 AGGAAAACAAATATAGGTTTGAATCATGGAGAAAGAAATCGCTTTTAAATATGA 3540 QY  
3481 AGGAAAACAAATATAGGTTTGAATCATGGAGAAAGAAATCGCTTTTAAATATGA 3540 Db  
3541 TTGCTACCAAGTCTTCTCATGTTGGTGGCAGTTGGATGATGATGATGATGATGATGATGATG 3600 QY  
3541 TTGCTACCAAGTCTTCTCATGTTGGTGGCAGTTGGATGATGATGATGATGATGATGATGATG 3600 Db  
3601 AGTTTGTTCGCTGAATGACAACTGACCAATCATTAAGATGCTCAGACAGTGAAGG 3660 QY  
3601 AGTTTGTTCGCTGAATGACAACTGACCAATCATTAAGATGCTCAGACAGTGAAGG 3660 Db  
3661 CTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCAATACCTTCCCAATTTGAACTGCCAA 3720 QY  
3661 CTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCAATACCTTCCCAATTTGAACTGCCAA 3720 Db  
3721 GAGAGTATCGAAACGTTTCTTCATATGATGATGATGATGATGATGATGATGATGATGATG 3780 QY  
3721 GAGAGTATCGAAACGTTTCTTCATATGATGATGATGATGATGATGATGATGATGATGATG 3780 Db  
3781 ACAAAATGAAATTTGGAGCCCAATTTGATGATGATGATGATGATGATGATGATGATGATG 3840 QY  
3781 ACAAAATGAAATTTGGAGCCCAATTTGATGATGATGATGATGATGATGATGATGATGATG 3840 Db  
3841 CATTTTCTGAGCAGTTTAAATGCACTTAAAGCGGAAGATATTTCCAGAGAGGAGATAC 3900 QY  
3841 CATTTTCTGAGCAGTTTAAATGCACTTAAAGCGGAAGATATTTCCAGAGAGGAGATAC 3900 Db  
3901 ACAAGAGCTAGTCCCAATCGAATCAGATGATGATGATGATGATGATGATGATGATGATG 3960 QY  
3901 ACAAGAGCTAGTCCCAATCGAATCAGATGATGATGATGATGATGATGATGATGATGATG 3960 Db  
3961 CCTCAGCATTTACTGAGCATTTTAAACTCAGCTTCAAGAGATGCTTTTGTGATGTGAT 4020 QY  
3961 CCTCAGCATTTACTGAGCATTTTAAACTCAGCTTCAAGAGATGCTTTTGTGATGTGAT 4020 Db  
4021 GCTTAGCAGTTTGGCCCGAAGAGAAATATCCAGTACCATGCTGTTTGTGTCATGAA 4080 QY

Db 4021 GCTTTAGCAGTGTGGCCGGAAGGAAAATATCCAGTACCATGCTGTGTTGTGGCATGAA 4080  
Qy 4081 TATAGCCCACTGACTAGGAATATTTAAACCAACCCTGAAACCTGTGTGTGCGAGCAGC 4140  
Db 4081 TATAGCCCACTGACTAGGAATATTTAAACCAACCCTGAAACCTGTGTGTGCGAGCAGC 4140  
Qy 4141 TCTGAACCTGATTTTACCTTTTAAAGAAATTTGCTCATGAGCCTGTCACTCTTTTATAAAAA 4200  
Db 4141 TCTGAACCTGATTTTACCTTTTAAAGAAATTTGCTCATGAGCCTGTCACTCTTTTATAAAAA 4200  
Qy 4201 GGCTCACTGACAAGACAGCTGTTAAATTTCCCAACCAATTCATGTCAGACTAACTTTAT 4260  
Db 4201 GGCTCACTGACAAGACAGCTGTTAAATTTCCCAACCAATTCATGTCAGACTAACTTTAT 4260  
Qy 4261 TAGGAGAAGCCTATGCCAGCTGGAGTGATTTGCTTAAGAGGCTCCAGCTTTTCAATCCAA 4320  
Db 4261 TAGGAGAAGCCTATGCCAGCTGGAGTGATTTGCTTAAGAGGCTCCAGCTTTTCAATCCAA 4320  
Qy 4321 AGCCTTTTGTAAAGTTTGCACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 4380  
Db 4321 AGCCTTTTGTAAAGTTTGCACCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 4380  
Qy 4381 AGTTAACTAGTTATCTTTGCTTCTGAGTATAACGAATTTGGGATGCTCTAAACCTATTTT 4440  
Db 4381 AGTTAACTAGTTATCTTTGCTTCTGAGTATAACGAATTTGGGATGCTCTAAACCTATTTT 4440  
Qy 4441 TAGATGTTTATTAATTAATGAGCAATATCACCTCTTATTTGACAACTAAATTTATGAG 4500  
Db 4441 TAGATGTTTATTAATTAATGAGCAATATCACCTCTTATTTGACAACTAAATTTATGAG 4500  
Qy 4501 TTTTATTAATTTAAGACTGTAAATGGTCTTAAACCACTAACTACTGAAAGAGCTCAATG 4560  
Db 4501 TTTTATTAATTTAAGACTGTAAATGGTCTTAAACCACTAACTACTGAAAGAGCTCAATG 4560  
Qy 4561 ATTGACATCTGAATGCTTTGTAATTAATTTGACTTTCAGCCCTAAGAAAGCTATGATTTCA 4620  
Db 4561 ATTGACATCTGAATGCTTTGTAATTAATTTGACTTTCAGCCCTAAGAAAGCTATGATTTCA 4620  
Qy 4621 CGTGCAGCTCTAAATTTCAACAGGCTGAGTTAGTACTTACCTACCTGAGATGAAATATGTTT 4680  
Db 4621 CGTGCAGCTCTAAATTTCAACAGGCTGAGTTAGTACTTACCTACCTGAGATGAAATATGTTT 4680  
Qy 4681 TGGAAATGTACATATTTCAACAGAGTGCCTCAATTTTGAAGAACTGAGTGTCTGATGGCA 4740  
Db 4681 TGGAAATGTACATATTTCAACAGAGTGCCTCAATTTTGAAGAACTGAGTGTCTGATGGCA 4740  
Qy 4741 CTGGCACAATTACAGTGTGCTGTTTAACTACTCACTCACTGGTATATTCAGTACTACTCT 4800  
Db 4741 CTGGCACAATTACAGTGTGCTGTTTAACTACTCACTCACTGGTATATTCAGTACTACTCT 4800  
Qy 4801 CTCAGTTGGTTTTGTATAGAACAGAGCCAGCAAACTTTCTTTGTAAAGGCTGGTTAGT 4860  
Db 4801 CTCAGTTGGTTTTGTATAGAACAGAGCCAGCAAACTTTCTTTGTAAAGGCTGGTTAGT 4860  
Qy 4861 AAATTTATGAGCCACCTGTGCTTTGTATACATTTCTTTCTGCTGCTTTGTTAGTTGT 4920  
Db 4861 AAATTTATGAGCCACCTGTGCTTTGTATACATTTCTTTCTGCTGCTTTGTTAGTTGT 4920  
Qy 4921 TTTTTCACCAACACCTCTAATAATCTAATAATCTAATAATCTAATAATCTAATAATCTAATAA 4980  
Db 4921 TTTTTCACCAACACCTCTAATAATCTAATAATCTAATAATCTAATAATCTAATAATCTAATAA 4980  
Qy 4981 ACTGCCCAACCAAGAGTGTGACCTTCAGGCCATCAATTTGCCAATCACTGAGAATTTAT 5040  
Db 4981 ACTGCCCAACCAAGAGTGTGACCTTCAGGCCATCAATTTGCCAATCACTGAGAATTTAT 5040  
Qy 5041 TTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 5100  
Db 5041 TTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 5100  
Qy 5101 TGCAGTGGCGCAATCTCAGCTCACTGCAACCTCCGCTCCCGGTTTCAAGCAGTCTGTGC 5160

Db 5101 TCAGTGGCGCAATCTCAGCTCACTGCAACCTCCGCTCCCGGTTTCAAGCAGTCTGTGC 5160  
Qy 5161 TCAGCTCTTCTGAGTAGCTGGGACTACAGGTGATGCCACCACTGCTTAATTTTGTGA 5220  
Db 5161 TCAGCTCTTCTGAGTAGCTGGGACTACAGGTGATGCCACCACTGCTTAATTTTGTGA 5220  
Qy 5221 TTTTATGATAGAGACGGGGTTCACCATATTTGGTCAGGCTTATCTTGAACCTCTGACCTC 5280  
Db 5221 TTTTATGATAGAGACGGGGTTCACCATATTTGGTCAGGCTTATCTTGAACCTCTGACCTC 5280  
Qy 5281 AGGTGATCCACCTGCTCTGCTCCCAAGTGTGAGATTTACAGGCATAAAGCAGTGCAC 5340  
Db 5281 AGGTGATCCACCTGCTCTGCTCCCAAGTGTGAGATTTACAGGCATAAAGCAGTGCAC 5340  
Qy 5341 CCAGCCGAGAAATAGTATTTTATGATGTTAAACCTTGGGCTAGCCATATTTATG 5400  
Db 5341 CCAGCCGAGAAATAGTATTTTATGATGTTAAACCTTGGGCTAGCCATATTTATG 5400  
Qy 5401 TCATTAATCAATGGATTTGTGAAGAGCAGATTTCCATGATGTAACCTGTCAGCAGTATTTAG 5460  
Db 5401 TCATTAATCAATGGATTTGTGAAGAGCAGATTTCCATGATGTAACCTGTCAGCAGTATTTAG 5460  
Qy 5461 ATCATGATCTCAACAATATTTCTCCAAATGGCATACATCTTTTGTACAAAGAACTTGAA 5520  
Db 5461 ATCATGATCTCAACAATATTTCTCCAAATGGCATACATCTTTTGTACAAAGAACTTGAA 5520  
Qy 5521 ATGTAATACCTGTTTGTGCTGTAAGAGTGTGTAATTTCAAAAAGTGAATCTCATAAA 5580  
Db 5521 ATGTAATACCTGTTTGTGCTGTAAGAGTGTGTAATTTCAAAAAGTGAATCTCATAAA 5580  
Qy 5581 AAGTTAAATTTTGA AAAA 5597  
Db 5581 AAGTTAAATTTTGA AAAA 5597

RESULT 9  
US-09-895-072-20  
; Sequence 20, Application US/09895072  
; Patent No. US2002002550A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM M  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 210119USOCNT  
; CURRENT APPLICATION NUMBER: US/09/895,072  
; CURRENT FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; PRIOR APPLICATION NUMBER: US 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 20  
; LENGTH: 3783  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-09-895-072-20

Query Match 67.2%; Score 3761; DB 9; Length 3783;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 3767; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 165 ATGCTGTTCAAGCTCTGCGAGAGCAAAACCTTATACCTGCTGCTCCACAGGTATGGGCTC 224  
Db 7 ATGGGGTTCAAGCTCTTGGAGAGCAAAACCTTATACCTGCTGCTCCACAGGTATGGGCTC 66  
Qy 225 TACGTGTGCTTCTTGGGGTCTGTCACCATGCTCTCCGCTTCCAGTTCGGAGAGGTG 284  
Db 67 TACGTGTGCTTCTTGGGGTCTGTCACCATGCTCTCCGCTTCCAGTTCGGAGAGGTG 126  
Qy 285 GTTCTGGAATGAGCCGAGATCAATACCATGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTT 344  
Db 127 GTTCTGGAATGAGCCGAGATCAATACCATGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTT 186

QY 345 GCTGGAAGTCTTTTTCAGAAATCGGCTTTGTCTGCCATGCCGATTCGACGTTGTTTACACC 404  
Db 187 GCTGGAAGTCTTTTTCAGAAATCGGCTTTGTCTGCCATGCCGATTCGACGTTGTTTACACC 246  
QY 405 TGGGTGAATGGCAGAGATCTTGAACCTAAGGAACTACAGCAGGTGAGAGAACAGATG 464  
Db 247 TGGGTGAATGGCAGAGATCTTGAACCTAAGGAACTACAGCAGGTGAGAGAACAGATG 306  
QY 465 GAGGAGGAGCAGAGCAATGAGAGAAATCTTGGGAAACCAACGAGCACTACTAAG 524  
Db 307 GAGGAGGAGCAGAGCAATGAGAGAAATCTTGGGAAACCAACGAGCACTACTAAG 366  
QY 525 AAGAGTGAGAGCAGATGTAGAGTGTCTTAACACACTGCAATTAAGGTGCCAATGCTTGTGA 584  
Db 367 AAGAGTGAGAGCAGATGTAGAGTGTCTTAACACACTGCAATTAAGGTGCCAATGCTTGTG 426  
QY 585 CTGACCCAGCCCTGCCAGCAACATCACCTGAGAGAGCTGCCATCTCTTTATCCTTCT 644  
Db 427 CTGACCCAGCCCTGCCAGCAACATCACCTGAGAGAGCTGCCATCTCTTTATCCTTCT 486  
QY 645 TTTCAATCTGCCAGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGTC 704  
Db 487 TTTCAATCTGCCAGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGTC 546  
QY 705 TCAGTCTGTTTGTGACGATCTAAGAGTGTGAAGATGCCCACTCTGAGCTGCTTAAA 764  
Db 547 TCAGTCTGTTTGTGACGATCTAAGAGTGTGAAGATGCCCACTCTGAGCTGCTTAAA 606  
QY 765 GGAATAGCAGACAGATGAGGGGGTACTTGACACACATGAAGAGTCCCTGGA 824  
Db 607 GGAATAGCAGACAGATGAGGGGGTACTTGACACACATGAAGAGTCCCTGGA 666  
QY 825 TTAGTGCTAATGCAAGATTTGGCTTTCCTGAGTGGATTTCCACCAACATTCAGAGAAACA 884  
Db 667 TTAGTGCTAATGCAAGATTTGGCTTTCCTGAGTGGATTTCCACCAACATTCAGAGAAACA 726  
QY 885 AATCACTAAACCAAAATTTGCCAGAAATCTTTCTCTTAAGTCAAACTGTTGCAAGTTG 944  
Db 727 AATCACTAAACCAAAATTTGCCAGAAATCTTTCTCTTAAGTCAAACTGTTGCAAGTTG 786  
QY 945 TATTGAGGCGCAGTGTAGCGCTTTTAAACTGAATAACCCCAAGGATTTTCAAGAAATG 1004  
Db 787 TATTGAGGCGCAGTGTAGCGCTTTTAAACTGAATAACCCCAAGGATTTTCAAGAAATG 846  
QY 1005 AATGAGCAAACTAAGAAAGACATGACCAATGTAGTGAAGAACTGACCAATGATCTGCA 1064  
Db 847 AATGAGCAAACTAAGAAAGACATGACCAATGTAGTGAAGAACTGACCAATGATCTGCA 906  
QY 1065 TATTATTATGGGATCTGAGCGCATCAGCCAGTCTAAGCAGATGAAGACATCTCTGCC 1124  
Db 907 TATTATTATGGGATCTGAGCGCATCAGCCAGTCTAAGCAGATGAAGACATCTCTGCC 966  
QY 1125 AGTCGTTTGAAGATAACGAAGAACTGAGTACTCATTTGCGATCTATCGAGAGGCAATGCA 1184  
Db 967 AGTCGTTTGAAGATAACGAAGAACTGAGTACTCATTTGCGATCTATCGAGAGGCAATGCA 1026  
QY 1185 CCATGGGTTCCGGAATATTTTCAATGTCCACCAAGCGGAGATTCATCTCGGCTGGAACCTT 1244  
Db 1027 CCATGGGTTCCGGAATATTTTCAATGTCCACCAAGCGGAGATTCATCTCGGCTGGAACCTT 1086  
QY 1245 GACAACTCCTGAGTGACAAATAGTAACAACACAGGATGTTTTTCGAAATTTGAGCCACTTG 1304  
Db 1087 GACAACTCCTGAGTGACAAATAGTAACAACACAGGATGTTTTTCGAAATTTGAGCCACTTG 1146  
QY 1305 CCTACCTTTAGTTTCACTGCTATTTGAAGTCAATTCATCGCATCGAAGGCTGTCGCAAG 1364  
Db 1147 CCTACCTTTAGTTTCACTGCTATTTGAAGTCAATTCATCGCATCGAAGGCTGTCGCAAG 1206  
QY 1365 AAGTTTATTTACCTAAATGATGATGTCATGTTTGGGAAGGATGCTCTGGCCAGATGATTTT 1424  
Db 1207 AAGTTTATTTACCTAAATGATGATGTCATGTTTGGGAAGGATGCTCTGGCCAGATGATTTT 1266  
QY 1425 TACAGTCACTCCAAAGGCCAGAGGTTTATTGACATGGCCTGTGCCAACTGTGCCGAG 1484

Db 1267 TACAGTCACTCCAAAGGCCAGAGGTTTATTGACATGGCTGTGCCAACTGTGCCGAG 1326  
QY 1485 GGTGCCCCAGGTTTCTGGAATTAAGGATGGCTATTGTGCAAGGGCTTTAATAATTCAACC 1544  
Db 1327 GGTGCCCCAGGTTTCTGGAATTAAGGATGGCTATTGTGCAAGGGCTTTAATAATTCAACC 1386  
QY 1545 TGCAGATTGGGATGTGGGGATTGCTCTGGAAACAGTGGAGGAGTCCCTATATTGTCAGGA 1604  
Db 1387 TGCAGATTGGGATGTGGGGATTGCTCTGGAAACAGTGGAGGAGTCCCTATATTGTCAGGA 1446  
QY 1605 GGTGGAGGTACTGGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGTGGAGGAATAAAC 1664  
Db 1447 GGTGGAGGTACTGGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGTGGAGGAATAAAC 1506  
QY 1665 AGTGTCTCTTACTGTAATCAGGGATGTGCGAATTTCTGGCTCGCTGATATAAGTTCTGTGAC 1724  
Db 1507 AGTGTCTCTTACTGTAATCAGGGATGTGCGAATTTCTGGCTCGCTGATATAAGTTCTGTGAC 1566  
QY 1725 CAAGCATGCAATGTCTTGTCTGTTGGTGTGATGCTGGCAGTGTGGCAGAGATCATTTT 1784  
Db 1567 CAAGCATGCAATGTCTTGTCTGTTGGTGTGATGCTGGCAGTGTGGCAGAGATCATTTT 1626  
QY 1785 CATGAAATGTAATGAAGTGAATCTTCCCAACCCAGACTCACTATATATTTCGAAAGGT 1844  
Db 1627 CATGAAATGTAATGAAGTGAATCTTCCCAACCCAGACTCACTATATATTTCGAAAGGT 1686  
QY 1845 GAATGCTCTGCTTATTTTTCAGCTTTGCAAGAGTGTGCAAGAGGAGTGAAGGTGCTCAT 1904  
Db 1687 GAATGCTCTGCTTATTTTTCAGCTTTGCAAGAGTGTGCAAGAGGAGTGAAGGTGCTCAT 1746  
QY 1905 AGTGACAAATCCAATTAATTCGACATGCTTCTATTGCGAAACAGTGGAAACCCATCCACCTC 1964  
Db 1747 AGTGACAAATCCAATTAATTCGACATGCTTCTATTGCGAAACAGTGGAAACCCATCCACCTC 1806  
QY 1965 ATAATGACAGTGTGAATGATGCCACCAACATACATTTTAACTCTCAGTTTCAAAATACA 2024  
Db 1807 ATAATGACAGTGTGAATGATGCCACCAACATACATTTTAACTCTCAGTTTCAAAATACA 1866  
QY 2025 AACGATGAAGAGTTTCAAAATGCAAGTGTGAGTGTGACACAGAGGAGGACCAAAA 2084  
Db 1867 AACGATGAAGAGTTTCAAAATGCAAGTGTGAGTGTGACACAGAGGAGGACCAAAA 1926  
QY 2085 CTGAATTTCTACGCCCCAGAGGGTTACGAAATTTGTAGTCCCAATACACATCTTCTCCA 2144  
Db 1927 CTGAATTTCTACGCCCCAGAGGGTTACGAAATTTGTAGTCCCAATACACATCTTCTCCA 1986  
QY 2145 GAGGCGGAAATCTCTTTTGTAGGATATTCCCAAGAAACAGCTTCCCGAAGTTTAAGAGA 2204  
Db 1987 GAGGCGGAAATCTCTTTTGTAGGATATTCCCAAGAAACAGCTTCCCGAAGTTTAAGAGA 2046  
QY 2205 CATGATGTTAATCTCAACAGAGAGCCCGAGGAGGTGAAATTTCCCTCGGTAAATATT 2264  
Db 2047 CATGATGTTAATCTCAACAGAGAGCCCGAGGAGGTGAAATTTCCCTCGGTAAATATT 2106  
QY 2265 TCACCTCTCTCCAAAGACGCCAGTGTGAGTCTCAATACCTTTGATTTTGCACACTGAAACAT 2324  
Db 2107 TCACCTCTCTCCAAAGACGCCAGTGTGAGTCTCAATACCTTTGATTTTGCACACTGAAACAT 2166  
QY 2325 GGAGACATCACTTTTGAAGGATCAATTTGTGCAAGTCAAGCTTGTGAGATCAATTTCTG 2384  
Db 2167 GGAGACATCACTTTTGAAGGATCAATTTGTGCAAGTCAAGCTTGTGAGATCAATTTCTG 2226  
QY 2385 ATGAACCTCAGACATGCTTAAATTAATAATCAAGCTAATAACAGATGAACAAATGAC 2444  
Db 2227 ATGAACCTCAGACATGCTTAAATTAATAATCAAGCTAATAACAGATGAACAAATGAC 2286  
QY 2445 AGTTTGTGGTCCACAGGAAAAACAGGTTCAATAAAGCATTTTGGCCAAACAGCTTAGGA 2504  
Db 2287 AGTTTGTGGTCCACAGGAAAAACAGGTTCAATAAAGCATTTTGGCCAAACAGCTTAGGA 2346  
QY 2505 GTGTCTGAAGATTGACAGGTTGACATTTTCTGCTGAGTGTGAAAGTGAATGTTGTCAT 2564

Db 2347 GTGCTGAAAGATTGCAGAGGTTGACCTTTTCCTGCGAGTGAGTGTAAAGTGAATGGTTCAT 2406  
Qy 2565 GACCAGGGTCAGAAATCCACCCCTCGACTTGGAGACCAACAGCAAGATTAGAGTGGAAACT 2624  
Db 2407 GACCAGGGTCAGAAATCCACCCCTCGACTTGGAGACCAACAGCAAGATTAGAGTGGAAACT 2466  
Qy 2625 CACACCCCAAAACCATAGGCGGAAATGTGACAAAGAAAGCCCCCATCTCTGATTTGTT 2684  
Db 2467 CACACCCCAAAACCATAGGCGGAAATGTGACAAAGAAAGCCCCCATCTCTGATTTGTT 2526  
Qy 2685 CCACCTGGAAGCCAGATGACAAAGAAAGAAATACAGGGAAGAAAGAAAGAACAGT 2744  
Db 2527 CCACCTGGAAGCCAGATGACAAAGAAAGAAAGAAATACAGGGAAGAAAGAAAGAACAGT 2586  
Qy 2745 AGAATGAGGAAATGCTGAAATTCATATAGGCGTTTACTGAAGTGTTCATTGGAAGAAAG 2804  
Db 2587 AGAATGAGGAAATGCTGAAATTCATATAGGCGTTTACTGAAGTGTTCATTGGAAGAAAG 2646  
Qy 2805 CTGACGATATACAGATAGTTACTTGGCTTTTTCCTGCGGAGGAAAGAAAGTATTTTC 2864  
Db 2647 CTGACGATATACAGATAGTTACTTGGCTTTTTCCTGCGGAGGAAAGAAAGTATTTTC 2706  
Qy 2865 CAAGATCTTCTCGAGGAGAGAGTCAATGGAAGACAAATTTGGCATACTTCACTGATAGC 2924  
Db 2707 CTAGATCTTCTCGAGGAGAGAGTCAATGGAAGACAAATTTGGCATACTTCACTGATAGC 2766  
Qy 2925 AAAAATACTGGAGGCAACTAAAGATACATTTGCAGATTCCTCCAGATATGAAATAAA 2984  
Db 2767 AAGAAATACTGGAGGCAACTAAAGATACATTTGCAGATTCCTCCAGATATGAAATAAA 2826  
Qy 2985 ATTCTAAATAGCAAGTTGGATTCCATCGGGGAAAGTCCCTGCTCAGATGCTCAGATG 3044  
Db 2827 ATTCTAAATAGCAAGTTGGATTCCATCGGGGAAAGTCCCTGCTCAGATGCTCAGATG 2886  
Qy 3045 ATTGACCGGATGTTATGCAAGAACTGCAAGATATGTTCCCTGGAAGAAATTTGACAGAG 3104  
Db 2887 ATTGACCGGATGTTATGCAAGAACTGCAAGATATGTTCCCTGGAAGAAATTTGACAGAG 2946  
Qy 3105 TCATTTCACAAAGTGGCCATTCTGAGGATATGAGTTGCTTCTCTTATTTTAT 3164  
Db 2947 TCATTTCACAAAGTGGCCATTCTGAGGATATGAGTTGCTTCTCTTATTTTAT 3006  
Qy 3165 CTCATGATGTCAGTGCAGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGAT 3224  
Db 3007 CTCATGATGTCAGTGCAGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGAT 3066  
Qy 3225 CAATCTGGTGTCTGTCAGAGAAATCCGAACACTGGCTACCCAGAAATTCAGCACTG 3284  
Db 3067 CAATCTGGTGTCTGTCAGAGAAATCCGAACACTGGCTACCCAGAAATTCAGCACTG 3126  
Qy 3285 CCGTTAAGTTTGCAGGATTTGACAGGCTGGAACACATGCTTAATAAATTTGCTCAAAAATG 3344  
Db 3127 CCGTTAAGTTTGCAGGATTTGACAGGCTGGAACACATGCTTAATAAATTTGCTCAAAAATG 3186  
Qy 3345 CTTCTGCTGATATCAGCAGCTAAATAATATCCCAACTCAGGAAATCCTACTATGAT 3404  
Db 3187 CTTCTGCTGATATCAGCAGCTAAATAATATCCCAACTCAGGAAATCCTACTATGAT 3246  
Qy 3405 CCCAACCTGCCACCGTCACTAAAGTCTAGTACAACTGTAACCAAGTAACTGACAAA 3464  
Db 3247 CCCAACCTGCCACCGTCACTAAAGTCTAGTACAACTGTAACCAAGTAACTGACAAA 3306  
Qy 3465 ATCCAAAGCATATAAGGACAAACAAATATATAGTTTGAATTCATGGGAGAGAGAA 3524  
Db 3307 ATCCAAAGCATATAAGGACAAACAAATATATAGTTTGAATTCATGGGAGAGAGAA 3366  
Qy 3525 ATCCGTTTTAAATGATTCGACCAAGTTTCTGATGTTGGTGGCCAGTTGGATGACATA 3584  
Db 3367 ATCCGTTTTAAATGATTCGACCAAGTTTCTCATGTTGGTGGCCAGTTGGATGACATA 3426  
Qy 3585 AGAAAAACCCCTAGGAAGTTTGTGCTGATGACCAACATTTGACCAACATTCATAAAGAT 3644  
Db 3427 AGAAAAACCCCTAGGAAGTTTGTGCTGATGACCAACATTTGACCAACATTCATAAAGAT 3486

Qy 3645 GCTCAGACAGTGAAGCTGTTCTCAGGACCTTCTATGAATCCATGTTCCCATACCTTCC 3704  
Db 3487 GCTCAGACAGTGAAGCTGTTCTCAGGACCTTCTATGAATCCATGTTCCCATACCTTCC 3546  
Qy 3705 CAATTGTAAGTCCCAAGAGAGTATCGAAACCGTTTCTTCTATGATCATGATGATGAGGAA 3764  
Db 3547 CAATTGTAAGTCCCAAGAGAGTATCGAAACCGTTTCTTCTATGATCATGATGATGAGGAA 3606  
Qy 3765 TGGAGGCTTATCGAGACAAATTTGAAGTTTGGACCCATTTGTGTAAGCAACATTGATT 3824  
Db 3607 TGGAGGCTTATCGAGACAAATTTGAAGTTTGGACCCATTTGTGTAAGCAACATTGATT 3666  
Qy 3825 ATGTTTACTATATCTCATTTTCTGAGCAGTTAATGCACTTAAGCGGAGATATTT 3884  
Db 3667 ATGTTTACTATATCTCATTTTCTGAGCAGTTAATTTGCACTTTAAGCGGAGATATTT 3726  
Qy 3885 CCCAGAGGAGGATACACAAAGAGTATGTCCTCAATCGAATCAGAGTATAGAGATC 3941  
Db 3727 CCCAGAGGAGGATACACAAAGAGTATGTCCTCAATCGAATCAGAGTATAGAGATC 3783

RESULT 10  
US-09-986-552-20  
; Sequence 20, Application US/09986552  
; Patent No. US20030150981A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASE;  
; FILE REFERENCE: 215089US77DIV  
; CURRENT APPLICATION NUMBER: US/09/986,552  
; CURRENT FILING DATE: 2001-11-09  
; PRIOR APPLICATION NUMBER: 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 20  
; LENGTH: 3783  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-09-986-552-20

Query Match 67.2%; Score 3761; DB 9; Length 3783;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 3767; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 165 ATGCTGTTCAAGCTCCTGCAGAGACAAACCTATACCTGCTGCCCTCCAGGTATGGGCTC 224  
Db 7 ATGGGGTTCAAGCTCTTGACAGAGACAAACCTATACCTGCTGCCCTCCAGGTATGGGCTC 66  
Qy 225 TACGTGTGCTTCTTGGGGTCTGTTCTCACCCTCGTCTCGCTCCAGTTCGGAGAGTG 284  
Db 67 TACGTGTGCTTCTTGGGGTCTGTTCTCACCCTCGTCTCGCTCCAGTTCGGAGAGTG 126  
Qy 285 GTTCTGGAATGGAGCGAGATCAATACCATGTTTGTGTTTGTATTCCTATAGACAAATTT 344  
Db 127 GTTCTGGAATGGAGCGAGATCAATACCATGTTTGTGTTTGTATTCCTATAGACAAATTT 186  
Qy 345 GCTGGAAGTCTTTTCAGATCGGCTTGTCTGCCCATCGCGATTGACGTTGTTTACACC 404  
Db 187 GCTGGAAGTCTTTTCAGATCGGCTTGTCTGCCCATCGCGATTGACGTTGTTTACACC 246  
Qy 405 TGGGTCAATGGCACAGATCTTGAACTACTGAAGGAACCTACAGCAGGTTCAGAGAACAGATG 464  
Db 247 TGGGTCAATGGCACAGATCTTGAACTACTGAAGGAACCTACAGCAGGTTCAGAGAACAGATG 306  
Qy 465 GAGGAGGAGCAGAAAGCAATGAGAAATCCTTTGGGAAAAACACAAACGGAACCTACTAAG 524  
Db 307 GAGGAGGAGCAGAAAGCAATGAGAAATCCTTTGGGAAAAACACAAACGGAACCTACTAAG 366  
Qy 525 AAGAGTGAGAACAGTTAGAGTGTGTTGTTTAAACACACTGCAATTAAGTTCGCAATGCTTCTGA 584



367	Db		AAGAGT	GAGAAG	CAGT	TTAGAG	TGTTT	GCTTAAC	ACACAC	TGCATTAAG	GTGCCAA	TGCTT	GC	426
585	Qy		CTGAC	CCAGC	CGCTG	CCAGC	CAACAT	CACCT	GGAAG	CAGTGC	CACTCT	TTTAT	CTCT	644
427	Db		CTGAC	CCAGC	CGCTG	CCAGC	CAACAT	CACCT	GGAAG	CACTGC	CACTCT	TTTAT	CTCT	486
645	Qy		TTTCAT	TCTG	CAGT	GCACAT	TTTCAAT	TGTTG	CAAAAC	CAAAAC	CCCTT	CTACCA	ATGC	704
487	Db		TTTCAT	TCTG	CAGT	GCACAT	TTTCAAT	TGTTG	CAAAAC	CAAAAC	CCCTT	CTACCA	ATGC	546
705	Qy		TCAGT	TGTTG	TTTGA	CAGTACT	TAAGAT	TGTTGA	AGATGCC	CACTCTG	GAC	TGCTT	AAA	764
547	Db		TCAGT	TGTTG	TTTGA	CAGTACT	TAAGAT	TGTTGA	AGATGCC	CACTCTG	GAC	TGCTT	AAA	606
765	Qy		GGAAT	TAGC	AGAC	AGATG	TAGGG	GGGTACT	TGCAAC	AGATAA	GAAGTCC	CTGG	824	
607	Db		GGAAT	TAGC	AGAC	AGATG	TAGGG	GGGTACT	TGCAAC	AGATAA	GAAGTCC	CTGG	666	
825	Qy		TTAGT	GTCT	TAATG	CAAGAT	TTTGGCTT	CTGAG	TGGATTTCC	ACCAACAT	TCAAG	GAACA	884	
667	Db		TTAGT	GTCT	TAATG	CAAGAT	TTTGGCTT	CTGAG	TGGATTTCC	ACCAACAT	TCAAG	GAACA	726	
885	Qy		AATCA	ACTT	AAAAAC	CAAAAT	TGCCA	AGAAATCT	TTTCC	TAAAGT	CAAACTG	TGTCAG	944	
727	Db		AATCA	ACTT	AAAAAC	CAAAAT	TGCCA	AGAAATCT	TTTCC	TAAAGT	CAAACTG	TGTCAG	786	
945	Qy		TATTC	AGAG	CCAGT	GTAG	CGCTTCT	TAAAACT	GAATAA	CCCCA	AGGATTTT	CAAGA	1004	
787	Db		TATTC	AGAG	CCAGT	GTAG	CGCTTCT	TAAAACT	GAATAA	CCCCA	AGGATTTT	CAAGA	846	
1005	Qy		AATA	AGC	AAAACT	TAAG	AGAAAC	TGAC	CAAT	TGTAAG	GAAGAA	CACTG	1064	
847	Db		AATA	AGC	AAAACT	TAAG	AGAAAC	TGAC	CAAT	TGTAAG	GAAGAA	CACTG	906	
1065	Qy		TATTT	ATTAT	TGGAT	CTG	AGC	GGCCAT	CAGC	CAGTCTA	AGCAGG	ATGA	1124	
907	Db		TATTT	ATTAT	TGGAT	CTG	AGC	GGCCAT	CAGC	CAGTCTA	AGCAGG	ATGA	966	
1125	Qy		AGTC	GTTTT	TGA	AGTA	ACG	AGAACT	CAGAGT	CACTCAT	TGGATCT	TCG	1184	
967	Db		AGTC	GTTTT	TGA	AGTA	ACG	AGAACT	CAGAGT	CACTCAT	TGGATCT	TCG	1026	
1185	Qy		CCAT	TGGTT	TCG	GAATAT	TTTT	CATT	TGTCAC	CAACGG	GCAGAT	TCAT	1244	
1027	Db		CCAT	TGGTT	TCG	GAATAT	TTTT	CATT	TGTCAC	CAACGG	GCAGAT	TCAT	1086	
1245	Qy		GAC	AACT	CTG	AGTGA	CAAT	AGT	TAACAC	ACAGAG	TGTTTT	TG	1304	
1087	Db		GAC	AACT	CTG	AGTGA	CAAT	AGT	TAACAC	ACAGAG	TGTTTT	TG	1146	
1305	Qy		CTTAC	CTTT	TAGT	TC	CACCT	CGT	TATTG	AAAGTCA	ATTCAT	CGCAT	1364	
1147	Db		CTTAC	CTTT	TAGT	TC	CACCT	CGT	TATTG	AAAGTCA	ATTCAT	CGCAT	1206	
1365	Qy		AAGT	TTTAT	TACCT	TAAT	ATGAT	GTGCAT	GT	TGGGA	AGGATGT	TG	1424	
1207	Db		AAGT	TTTAT	TACCT	TAAT	ATGAT	GTGCAT	GT	TGGGA	AGGATGT	TG	1266	
1425	Qy		TAC	AGT	CAC	TCC	AAAG	CC	AGAGT	TTTATT	TGCA	TG	1484	
1267	Db		TAC	AGT	CAC	TCC	AAAG	CC	AGAGT	TTTATT	TGCA	TG	1326	
1485	Qy		GGCT	GCC	CAG	GTTC	CTG	GAAT	AGGAT	TGCTAT	TGCA	AA	1544	
1327	Db		GGCT	GCC	CAG	GTTC	CTG	GAAT	AGGAT	TGCTAT	TGCA	AA	1386	
1545	Qy		TGCG	ATT	TGG	AT	TG	TGG	GA	AA	CAGT	TG	1604	
1387	Db		TGCG	ATT	TGG	AT	TG	TGG	GA	AA	CAGT	TG	1446	
1605	Qy		GGT	GG	AG	GT	TACT	TGG	AGT	TG	GA	CC	1664	

D	b	1447	GGTGGAGGTACTCGGAGTA'TTGAGTT'GGACAGCCCTCGCAGTTTGGTGAGGAATAAAC	1506
Q	y	1665	AGTGTCTTACTACTGAATCAAGGATGTCGAAATT'CCTGGCTCGCTGATAGTTCTCTGTAC	1724
D	b	1507	AGTGTCTTACTACTGAATCAAGGATGCGAAATTCCTGGCTCGCTGATAGTTCTCTGTAC	1566
Q	y	1725	CAAGCATGCAAATGTCCTTGTCTCTGTGGGTTTGATGCTGGCGACTGTGGCGAAGATCAATTT	1784
D	b	1567	CAAGCATGCATGTCCTTGTCTCTGTGGGTTTGATGCTGGCGACTGTGGCGAAGATCAATTT	1626
Q	y	1785	CATGAATTTGATAAAGTGATCTCTTCTCCCAAACAGACTCACTATATTTATTTCCAAAAGGT	1844
D	b	1627	CATGAATTTGATAAAGTGATCTCTTCTCCCAAACAGACTCACTATATTTATTTCCAAAAGGT	1686
Q	y	1845	GAATGCCCTGCTTATTTTTCAGCTTTGCGAGAAGTAGCCAAAAAGAGAGGTTGAAGTGGCTAT	1904
D	b	1687	GAATGCCCTGCTTATTTTTCAGCTTTGCGAGAAGTAGCCAAAAAGAGAGGTTGAAGTGGCTAT	1746
Q	y	1905	AGTGACAAATCCAATAATTCGACATGCTTCTATTGCCCACAAGTGGAAAAACCATCCACCTC	1964
D	b	1747	AGTGACAAATCCAATAATTCGACATGCTTCTATTGCCCACAAGTGGAAAAACCATCCACCTC	1806
Q	y	1965	ATAATGCACAGTGGAAATGAATGCCACCACCAATACATTTTAAATCTCACGTTTCCAAAATACA	2024
D	b	1807	ATAATGCACAGTGGAAATGAATGCCACCACCAATACATTTTAAATCTCACGTTTCCAAAATACA	1866
Q	y	2025	AACGATGAGAGTTCAAANNTGCAGATACACAGTGGAGTGGACACACAGGGGAGGCACAAA	2084
D	b	1867	AACGATGAGAGTTCAAANNTGCAGATACACAGTGGAGTGGACACACAGGGGAGGCACAAA	1926
Q	y	2085	CTGAATTTCTACGGCCCCAGAGGGTTACGAAAAATTTAGTTAGTCCCATAAACATTTCTTCCA	2144
D	b	1927	CTGAATTTCTACGGCCCCAGAGGGTTACGAAAAATTTAGTTAGTCCCATAAACATTTCTTCCA	1986
Q	y	2145	GAGCGGAAATCTTTTTTGAGGATATTTCCCAAAGAAAAACGCTTCCGGAAGTTTAAGAGA	2204
D	b	1987	GAGCGGAAATCTTTTTTGAGGATATTTCCCAAAGAAAAACGCTTCCGGAAGTTTAAGAGA	2046
Q	y	2205	CATGATGTTTAACTCAAACAAGAGAGCCAGAGAGGTGAAAAATTCCTCTGGTAAATATTT	2264
D	b	2047	CATGATGTTTAACTCAAACAAGAGAGCCAGAGAGGTGAAAAATTCCTCTGGTAAATATTT	2106
Q	y	2265	TCACTCTCTCCAAAAGAGCCGATTTGAGTCTCAATACCTTTGGATTTGCAATCTGGAACAT	2324
D	b	2107	TCACTCTCTCCAAAAGAGCCGATTTGAGTCTCAATACCTTTGGATTTGCAATCTGGAACAT	2166
Q	y	2325	GGACACATCACTTTGAAAGGATACAATTTCTGCAAGTCAGCCPTTGTGTAGATCATTTCTG	2384
D	b	2167	GGACACATCACTTTGAAAGGATACAATTTGTCCAGTCAAGCTTGTGTAGATCATTTCTG	2226
Q	y	2385	ATGAACTCACAGCATGCTTAAATAAATAATCAAGCTATATAACAGATGAAACAAATGAC	2444
D	b	2227	ATGAACTCACAGCATGCTTAAATAAATAATCAAGCTATATAACAGATGAAACAAATGAC	2286
Q	y	2445	AGTTTGTGGTCCACAGGAAAAACAGGTTTATAAAGCATCTTTCGCAACACCTTAGGA	2504
D	b	2287	AGTTTGTGGTCCACAGGAAAAACAGGTTTATAAAGCATCTTTCGCAACACACCTTAGGA	2346
Q	y	2505	GTGTCTGAAAGATTGACAGAGTTGACATTTTCTCTGCAGTGAAGTGAATGGTTCAT	2564
D	b	2347	GTGTCTGAAAGATTGACAGAGTTGACATTTTCTCTGCAGTGAAGTGAATGGTTCAT	2406
Q	y	2565	GACAGGGTCAGAATCCACCCCTGGACTTTGGAGNCCACAGCAAGATTTAGAGTGGAACT	2624
D	b	2407	GACAGGGTCAGAATCCACCCCTGGACTTTGGAGNCCACAGCAAGATTTAGAGTGGAACT	2466
Q	y	2625	CACACCCAAAAACCATAGCGGGAATGTGCACAAAAAGAAAGCCCCCATCTCTGATTTGTT	2684
D	b	2467	CACACCCAAAAACCATAGCGGGAATGTGCACAAAAAGAAAGCCCCCATCTCTGATTTGTT	2526
Q	y	2685	CCACTGNAAGCCAGATTGACAAAGAAAAAGAAATACAGGGGAAAGAAAAAGAGAACAGT	2744
D	b	2527	CCACTGNAAGCCAGATTGACAAAGAAAAAGAAATACAGGGGAAAGAAAAAGAGAACAGT	2586



QY	2745	AGAAATGGAGGGAANAATGCTGAAAAATCA	CATAGGCCGTTACTGAAAGTGTTACTTTGGAAGAAG	2804
DB	2587	AGAAATGGAGGGAANAATGCTGAAAAATCA	CATAGGCCGTTACTGAAAGTGTTACTTTGGAAGAAG	2646
QY	2805	CTGCAGCATTACACAGATAGT	TACTTGGCGCTTTTGGCATGGGAGAAAAAAGTATTTC	2864
DB	2647	CTGCAGCATTACACAGATAGT	TACTTGGCGCTTTTGGCATGGGAGAAAAAAGTATTTC	2706
QY	2865	CAAGATCTTCTCGAGGAGAGAGAGT	CATTGAAACACAAATGGGCATCTTCACCTGATAGC	2924
DB	2707	CTAGATCTTCTCGAGGAGAGAGAGT	CATTGAAACACAAATGGGCATCTTCACCTGATAGC	2766
QY	2925	AAAAATACTGGGAGGCAACTAAAAAGATAC	ATTTCAGATTCCCTCAGATATGTAATAAAA	2984
DB	2767	AAGAAATACTGGGAGGCAACTAAAAAGATAC	ATTTCAGATTCCCTCAGATATGTAATAAAA	2826
QY	2985	ATTCTAAATAGCAGATTGGAATTCACATCG	GGGAAAGTCCCTGCTCACAATGCCTCACATG	3044
DB	2827	ATTCTAAATAGCAGATTGGAATTCACATCG	GGGAAAGTCCCTGCTCACAATGCCTCACATG	2886
QY	3045	ATTGACCGGATTTGTTATGCAAGAACTGCA	AGATATGTTCCCTGAAGAAATTTGACAAGACG	3104
DB	2887	ATTGACCGGATTTGTTATGCAAGAACTGCA	AGATATGTTCCCTGAAGAAATTTGACAAGACG	2946
QY	3105	TCATTTCAAAAGTCGCGCAATCTGAGGATAT	GAGTATGCGCTTCTCTTATTTTTTATAT	3164
DB	2947	TCATTTCAAAAGTCGCGCAATCTGAGGATAT	GAGTATGCGCTTCTCTTATTTTTTATAT	3006
QY	3165	CTCATGAGTGCAGTGCAGCCACTCAATATAT	CTCAAGTCTTTGATGAAGTTGATACAGAT	3224
DB	3007	CTCATGAGTGCAGTGCAGCCACTCAATATAT	CTCAAGTCTTTGATGAAGTTGATACAGAT	3066
QY	3225	CAATCTCGTGTCTGTCGACAGAGAAATCG	GAACACATGGCTACCGAAATTCACGAACTG	3284
DB	3067	CAATCTCGTGTCTGTCGACAGAGAAATCG	GAACACATGGCTACCGAAATTCACGAACTG	3126
QY	3285	CCGTTAAGTTTGACAGGATTTGACAGGTC	TGGAACACATGCTTAATAATTTGCTCAAAAATG	3344
DB	3127	CCGTTAAGTTTGACAGGATTTGACAGGTC	TGGAACACATGCTTAATAATTTGCTCAAAAATG	3186
QY	3345	CTTCCTGCTGATATCACGAGCTAAATAAT	TCCACCAACTCAGGAAATCCTACTATGAT	3404
DB	3187	CTTCCTGCTGATATCACGAGCTAAATAAT	TCCACCAACTCAGGAAATCCTACTATGAT	3246
QY	3405	CCCAACCTGCCACCGGTCTCTAAGGTC	TAGTAAACAACCTGTAACCAAGTAACCTGACAAA	3464
DB	3247	CCCAACCTGCCACCGGTCTCTAAGGTC	TAGTAAACAACCTGTAACCAAGTAACCTGACAAA	3306
QY	3465	ATCCACAAAGCATATAAGGCAAAAAA	CAAAATATAGGTTTGAAATCATGGGAGAGAAAGAA	3524
DB	3307	ATCCACAAAGCATATAAGGCAAAAAA	CAAAATATAGGTTTGAAATCATGGGAGAGAAAGAA	3366
QY	3525	ATCGCTTTTAAATGATTCTGACCAACG	TTTCTCATGTGTTGGCCAGTTGGATGACATA	3584
DB	3367	ATCGCTTTTAAATGATTCTGACCAACG	TTTCTCATGTGTTGGCCAGTTGGATGACATA	3426
QY	3585	AGAAAAACCTTAGGAAAGTTTGTGGCTGA	ATGACCAACATGACCAACATCATATAAGAT	3644
DB	3427	AGAAAAACCTTAGGAAAGTTTGTGGCTGA	ATGACCAACATGACCAACATCATATAAGAT	3486
QY	3645	GCTCAGACAGTGAAGGCTGTTCTCAGGGA	CTTCTATGAATCCATGTTTCCCGATACCTTCC	3704
DB	3487	GCTCAGACAGTGAAGGCTGTTCTCAGGGA	CTTCTATGAATCCATGTTTCCCGATACCTTCC	3546
QY	3705	CAATTTGAACTGCCAAGAGAGTATCGAAA	CCGTTTCTCTCATATGCAATGAGCTGCAGGAA	3764
DB	3547	CAATTTGAACTGCCAAGAGAGTATCGAAA	CCGTTTCTCTCATATGCAATGAGCTGCAGGAA	3606
QY	3765	TGGAGGCGCTTATCGAGACAAATGGAAG	TTTTGGACCCATTTGTGTACTAGCAACATTTGATT	3824
DB	3607	TGGAGGCGCTTATCGAGACAAATGGAAG	TTTTGGACCCATTTGTGTACTAGCAACATTTGATT	3666

Qy	3825	ATGTTTACTATATCTCATTTTCTGCTGACGACTTAATTCGCACTTAAGCGGAAGATATTT	3884
Db	3667	ATGTTTACTATATCTCATTTTCTGCTGACGACTTAATTCGCACTTAAGCGGAAGATATTT <td>3726</td>	3726
Qy	3885	CCCAGAAGGAGGATACACAAAGAAGCTAGTCCCAATCGAATCAGAGTATAGAAGATC	3941
Db	3727	CCCAGAAGGAGGATACACAAAGAAGCTAGTCCCAATCGAATCAGAGTATAGAAGATC	3783
RESULT 11			
US-10-306-686-20			
; Sequence 20, Application US/10306686			
; Publication No. US20030148460A1			
; GENERAL INFORMATION:			
; APPLICANT: CANFIELD, WILLIAM			
; TITLE OF INVENTION: PHOSPHOLIPID ALPHAS-GLUCOSIDASE OF THE LYSOSOMAL TARGETING			
; FILE REFERENCE: 230397US77DIV			
; CURRENT APPLICATION NUMBER: US/10/306,686			
; CURRENT FILING DATE: 2002-11-29			
; PRIOR APPLICATION NUMBER: 09/636,596			
; PRIOR FILING DATE: 2000-08-10			
; PRIOR APPLICATION NUMBER: 60/153,831			
; PRIOR FILING DATE: 1999-08-14			
; NUMBER OF SEQ ID NOS: 52			
; SOFTWARE: PatentIn version 3.1			
; SEQ ID NO 20			
; LENGTH: 3783			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-10-306-686-20			
Query Match 67.2%; Score 3761; DB 15; Length 3783;			
Best Local Similarity 99.74; Pred. No. 0;			
Matches 3767; Conservative 0; Mismatches 10; Indels 0; Gaps 0;			
Qy	165	ATGCTGTTCAGCTCCTGCGAGACAAACCTATACCTGCTGCTGCCACAGGTATGGGCTC	224
Db	7	ATGGGGTTCAGCTCTTTCGAGAGACAAACCTATACCTGCTGCTGCCACAGGTATGGGCTC	66
Qy	225	TAGTGTGCTCTTTGGCGCTGCTGTTCACATCGTCTCGCCCTTCAGTTCGAGAGGTG	284
Db	67	TACGTGTGCTCTTTGGCGCTGCTGTTCACATCGTCTCGCCCTTCAGTTCGAGAGGTG	126
Qy	285	GTTCTGGAATGGAGCGAGATCAATACCATTGTTTGTGTTGATTCCTATAGAGACAATATT	344
Db	127	GTTCCTGGAATGGAGCGAGATCAATACCATTGTTTGTGTTGATTCCTATAGAGACAATATT	186
Qy	345	GCTGGAAAGTCTTTTCAGAAATCGGCTTTGTCTGCCCATGCGGATTCGAGTGTGTTTACACC	404
Db	187	GCTGGAAAGTCTTTTCAGAAATCGGCTTTGTCTGCCCATGCGGATTCGAGTGTGTTTACACC	246
Qy	405	TGGTGTGATGGCACAGATCTTTGAATCTACTGAAGAACTACAGCAGGTTCAGAGAACAGATG	464
Db	247	TGGTGTGATGGCACAGATCTTTGAATCTACTGAAGAACTACAGCAGGTTCAGAGAACAGATG	306
Qy	465	GAGAGAGGACGAGAAAGCAATAGAGAAATCTCTGGGAAAAACAAACGGAACCTACTAAG	524
Db	307	GAGAGAGGACGAGAAAGCAATAGAGAAATCTCTGGGAAAAACAAACGGAACCTACTAAG	366
Qy	525	AAGAGTGAGAGCAGTATAGAGTCTTTGCTTAACACATGCAATTAAGGTGCCAATGCTTGA	584
Db	367	AAGAGTGAGAGCAGTATAGAGTCTTTGCTTAACACATGCAATTAAGGTGCCAATGCTTGA	426
Qy	585	CTGACCCAGCCCTGCGACCAACATCACCCCTGAAGGAGTGCATCTCTTTATCCTTCT	644
Db	427	CTGACCCAGCCCTGCGACCAACATCACCCCTGAAGGAGTGCATCTCTTTATCCTTCT	486
Qy	645	TTTTCATCTCGCAGTGACATTTCAATGTGCAAAACCAAAACCCCTTCTACCAATGTC	704
Db	487	TTTTCATCTCGCAGTGACATTTCAATGTGCAAAACCAAAACCCCTTCTACCAATGTC	546
Qy	705	TCAGTGTGTTGTTTTCGACAGTACTAAGAGTGTGAAGATGCCCACTCTGACATGCTTAAA	764

Db 547 TCAGTTGTTGTTTGTGACAGTACTAAGATGTTGAAGATGCCCACTCTCGACTGCTTAAA 606  
Qy 765 GGAATAAGCAGACAGACATGATGGAGGGTACTTGACAACAGATATAAGAACTCCCTGGA 824  
Db 607 GGAATAAGCAGACAGACATGATGGAGGGTACTTGACAACAGATATAAGAACTCCCTGGA 666  
Qy 825 TTAGTGCTAATGCAAGATTTGGCTTTCCCTGAGTGGATTTCCACCAACATTCAGGAAACA 884  
Db 667 TTAGTGCTAATGCAAGATTTGGCTTTCCCTGAGTGGATTTCCACCAACATTCAGGAAACA 726  
Qy 885 AATCAACTAAAAACAATAATGCGAGAAAATCTTTCTCTAAAGTCAAACTGTTGCAATTG 944  
Db 727 AATCAACTAAAAACAATAATGCGAGAAAATCTTTCTCTAAAGTCAAACTGTTGCAATTG 786  
Qy 945 TATTGAGGCCAGTGTAGCGCTTTCTAAACCTGAATAACCCCAAGGATTTTCAAGAAATTG 1004  
Db 787 TATTGAGGCCAGTGTAGCGCTTTCTAAACCTGAATAACCCCAAGGATTTTCAAGAAATTG 846  
Qy 1005 AATAAGCAAACTAAGAAAGAACATGACATGATGGAAGAAAGACTGACCAATAGTCTCTGCA 1064  
Db 847 AATAAGCAAACTAAGAAAGAACATGACATGATGGAAGAAAGACTGACCAATAGTCTCTGCA 906  
Qy 1065 TATTATTATGGGATCTGAGCGCCATCAGCCAGTCTAAGCAGATGAAGACATCTCTGCC 1124  
Db 907 TATTATTATGGGATCTGAGCGCCATCAGCCAGTCTAAGCAGATGAAGACATCTCTGCC 966  
Qy 1125 AGTCGTTTGAAGATRAACGAAGAACTGAGGTACTCATTCGGATCTATCGAGAGGCATGCA 1184  
Db 967 AGTCGTTTGAAGATRAACGAAGAACTGAGGTACTCATTCGGATCTATCGAGAGGCATGCA 1026  
Qy 1185 CCATGGTTGCGAATATTTTCATTGTGCAACCAAGCGCAGATTCCTCTGGCTGACCTT 1244  
Db 1027 CCATGGTTGCGAATATTTTCATTGTGCAACCAAGCGCAGATTCCTCTGGCTGACCTT 1086  
Qy 1245 GACAACTCCTGAGTGACAATAGTAACACACAGAGATGTTTTCGAAATTTGAGCCACTTG 1304  
Db 1087 GACAACTCCTGAGTGACAATAGTAACACACAGAGATGTTTTCGAAATTTGAGCCACTTG 1146  
Qy 1305 CTTACCTTTAGTTCACCTGCTATTGAAAGTCAATTCATCGAATCGAAGGCTGTCGCCAG 1364  
Db 1147 CTTACCTTTAGTTCACCTGCTATTGAAAGTCAATTCATCGAATCGAAGGCTGTCGCCAG 1206  
Qy 1365 AAGTTTATTTACCTTAAATGATGATGTCATGTTTGGGAAGATGCTGGCCAGATGATTTT 1424  
Db 1207 AAGTTTATTTACCTTAAATGATGATGTCATGTTTGGGAAGATGCTGGCCAGATGATTTT 1266  
Qy 1425 TACAGTCACTCCAAAGCCAGAGAGTTTATTGACATGGCCCTGTGCCAACTGTGCCGAG 1484  
Db 1267 TACAGTCACTCCAAAGCCAGAGAGTTTATTGACATGGCCCTGTGCCAACTGTGCCGAG 1326  
Qy 1485 GGCTGCCAGGTTCCCTGATTAAGATGGCTATTGTGACAAGGCTTGTAAATTCAGCC 1544  
Db 1327 GGCTGCCAGGTTCCCTGATTAAGATGGCTATTGTGACAAGGCTTGTAAATTCAGCC 1386  
Qy 1545 TGCATTTGGGATGGTGGGATTTGCTCTGGAACAGTGGAGGAGTCCGCTATATTGCGAG 1604  
Db 1387 TGCATTTGGGATGGTGGGATTTGCTCTGGAACAGTGGAGGAGTCCGCTATATTGCGAG 1446  
Qy 1605 GGTGGAGGTACTGGAGTATGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGATTAAC 1664  
Db 1447 GGTGGAGGTACTGGAGTATGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGATTAAC 1506  
Qy 1665 AGTGTCTCTTACTGTAATCAGGAGATGTCGAAATCTCTGGCTCGCTGATTAAGTTCTGTGAC 1724  
Db 1507 AGTGTCTCTTACTGTAATCAGGAGTATGGAGTTTGGTGGCTCGCTGATTAAGTTCTGTGAC 1566  
Qy 1725 CAAGCATGCAATGCTGTGCTGTGGGTTTGTGCTGGGCACTGTGGGCAAGATCATTTT 1784  
Db 1567 CAAGCATGCAATGCTGTGCTGTGGGTTTGTGCTGGGCACTGTGGGCAAGATCATTTT 1626  
Qy 1785 CATGAATTTGTAAGTGAATCTCTCTCCCAACAGACTCACTATATTATTCAAAAAGT 1844  
Db 1627 CATGAATTTGTAAGTGAATCTCTCTCCCAACAGACTCACTATATTATTCAAAAAGT 1686

Qy 1845 GAATGCTGCTGCTTATTTTTCAGCTTTGAGAAAGTAGCCAAAAAGAGAGTTGAAGTGGCTAT 1904  
Db 1687 GAATGCTGCTGCTTATTTTTCAGCTTTGAGAAAGTAGCCAAAAAGAGAGTTGAAGTGGCTAT 1746  
Qy 1905 AGTGACAAATCCAATTAATTTGCACATGCTTCTATTGCCCAACAAGTGGAAACCATCCACCTC 1964  
Db 1747 AGTGACAAATCCAATTAATTTGCACATGCTTCTATTGCCCAACAAGTGGAAACCATCCACCTC 1806  
Qy 1965 ATAATGCACAGTGGAAATGAATGCCACCAATACATTTTAATCTCAGCTTTCAAAATACA 2024  
Db 1807 ATAATGCACAGTGGAAATGAATGCCACCAATACATTTTAATCTCAGCTTTCAAAATACA 1866  
Qy 2025 AACGATGAAGAGTTCAAAATGCGAGATAACAGTGGAGTGGACACAAGGGAGGACCAAAA 2084  
Db 1867 AACGATGAAGAGTTCAAAATGCGAGATAACAGTGGAGTGGACACAAGGGAGGACCAAAA 1926  
Qy 2085 CTGAATTTCTAGCGCCCAAGAGGGTTACGAAAAATTTAGTTAGTCCCAATAACACTTTCTTCCA 2144  
Db 1927 CTGAATTTCTAGCGCCCAAGAGGGTTACGAAAAATTTAGTTAGTCCCAATAACACTTTCTTCCA 1986  
Qy 2145 GAGCGGAAATCTCTTTTGGAGATATTTCCCAAGAAAAACGCTTCCCGAAAGTTTAAGAGA 2204  
Db 1987 GAGCGGAAATCTCTTTTGGAGATATTTCCCAAGAAAAACGCTTCCCGAAAGTTTAAGAGA 2046  
Qy 2205 CATGATGTTAACTCAACAGAGAGCCCGAGAGAGGTGAAAAATTCCTCTGGTAAATATT 2264  
Db 2047 CATGATGTTAACTCAACAGAGAGCCCGAGAGAGGTGAAAAATTCCTCTGGTAAATATT 2106  
Qy 2265 TCACCTCTTCCAAAGACGCGCCAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAAACAT 2324  
Db 2107 TCACCTCTTCCAAAGACGCGCCAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAAACAT 2166  
Qy 2325 GGAGACATCACTTTGAAAGGATCAATTTGTGCAAGTCAGCCTTGTGAGATCAATTTCTG 2384  
Db 2167 GGAGACATCACTTTGAAAGGATCAATTTGTGCAAGTCAGCCTTGTGAGATCAATTTCTG 2226  
Qy 2385 ATGAACCTCAGACATGCTTAAATAAAAAATCAAGCTATAATAACAGATGAAAACAAATGAC 2444  
Db 2227 ATGAACCTCAGACATGCTTAAATAAAAAATCAAGCTATAATAACAGATGAAAACAAATGAC 2286  
Qy 2445 AGTTTGTGGTCTCACAAGGAAAAACAGGTTCAATAAAGCATCTTGGCAAAACAGCTTAGGA 2504  
Db 2287 AGTTTGTGGTCTCACAAGGAAAAACAGGTTCAATAAAGCATCTTGGCAAAACAGCTTAGGA 2346  
Qy 2505 GTGTCTGAAAGATTGACAGGTTGACCTTTCCCTGAGTGAAGTGAAGTGAATGGTCAAT 2564  
Db 2347 GTGTCTGAAAGATTGACAGGTTGACCTTTCCCTGAGTGAAGTGAAGTGAATGGTCAAT 2406  
Qy 2565 GACCAGGCTCAGAAATCCACCCCTGGACTTGGAGACCAACAGCAAGATTTAGAGTGGAAACT 2624  
Db 2407 GACCAGGCTCAGAAATCCACCCCTGGACTTGGAGACCAACAGCAAGATTTAGAGTGGAAACT 2466  
Qy 2625 CACACCCAAAAAACCATAGCGGAAATCTGACAAAAGAAAAGCCCATCTCTGATTTGTT 2684  
Db 2467 CACACCCAAAAAACCATAGCGGAAATCTGACAAAAGAAAAGCCCATCTCTGATTTGTT 2526  
Qy 2685 CCACCTGGAAAGCCAGATGACAAAAAGAAAAGAAAATCAAGGAAAGAAAAGAGAACAGT 2744  
Db 2527 CCACCTGGAAAGCCAGATGACAAAAAGAAAAGAAAATCAAGGAAAGAAAAGAGAACAGT 2586  
Qy 2745 AGAATGGAGGAAAAATGCTGAAAAATCAATAGCGCTTACTGAAGTGTACTTTGGAGAAAG 2804  
Db 2587 AGAATGGAGGAAAAATGCTGAAAAATCAATAGCGCTTACTGAAGTGTACTTTGGAGAAAG 2646  
Qy 2805 CTGACAGCAATTAACAGATAGTTACTTGGCTTTTTCATCGGAGAGAAAAAAGATATTTC 2864  
Db 2647 CTGACAGCAATTAACAGATAGTTACTTGGCTTTTTCATCGGAGAGAAAAAAGATATTTC 2706  
Qy 2865 CAAGATCTTCTCGACGAAGAGAGTCAATGAAGACAAATTTGGCATACTTCACTGATAGC 2924  
Db 2707 CTAGATCTTCTCGACGAAGAGAGTCAATGAAGACAAATTTGGCATACTTCACTGATAGC 2766

QY 2925 AAAAATCTGGAGGCACTAAAGATACATTTGCAGATTCCTCAGATATGTAATAAA 2984  
DB 2767 AAGAATCTGGAGGCACTAAAGATACATTTGCAGATTCCTCAGATATGTAATAAA 2826  
QY 2985 ATTCTAAATAGCAAGTTGGATTCACATCGCGGAAAGTCCCTGCTCAGATGCTCAGATG 3044  
DB 2827 ATTCTAAATAGCAAGTTGGATTCACATCGCGGAAAGTCCCTGCTCAGATGCTCAGATG 2886  
QY 3045 ATTGACCGGATTTGTTATGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGCAAGACG 3104  
DB 2887 ATTGACCGGATTTGTTATGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGCAAGACG 2946  
QY 3105 TCATTTCAAAGTGGCGCATTCGTAGGATATGAGTTGCGTTGCTTTCTTTATTTATTTAT 3164  
DB 2947 TCATTTCAAAGTGGCGCATTCGTAGGATATGAGTTGCGTTGCTTTCTTTATTTATTTAT 3006  
QY 3165 CTCATGAGTGCAGTGCAGCCACTCAATATATCTCAAGTCTTGTGATGAGTTGATACAGAT 3224  
DB 3007 CTCATGAGTGCAGTGCAGCCACTCAATATATCTCAAGTCTTGTGATGAGTTGATACAGAT 3066  
QY 3225 CAATCTGGTGTCTTGTCTGACAGAGAAATCCGAACTCTGGCTACAGAAATTCACGAACTG 3284  
DB 3067 CAATCTGGTGTCTTGTCTGACAGAGAAATCCGAACTCTGGCTACAGAAATTCACGAACTG 3126  
QY 3285 CCGTTAAGTTTGCAGGATTTGACAGGTTCTGGAACACATGCTTAATAATTGCTCAAAAATG 3344  
DB 3127 CCGTTAAGTTTGCAGGATTTGACAGGTTCTGGAACACATGCTTAATAATTGCTCAAAAATG 3186  
QY 3345 CTTCTCTGCTGATATCAGCGAGTAAATAATATCCACCAACTCAGGAATCCTACTATGAT 3404  
DB 3187 CTTCTCTGCTGATATCAGCGAGTAAATAATATCCACCAACTCAGGAATCCTACTATGAT 3246  
QY 3405 CCCAACTGCCACCGGTCTCTAAAGTCTAGTAAACAACTGTAAACACAGTAACTGACAAA 3464  
DB 3247 CCCAACTGCCACCGGTCTCTAAAGTCTAGTAAACAACTGTAAACACAGTAACTGACAAA 3306  
QY 3465 ATCCACAAAGCATATAGGACAAACAAATATAGGTTTCAATATCTAGGAGAGAGAA 3524  
DB 3307 ATCCACAAAGCATATAGGACAAACAAATATAGGTTTCAATATCTAGGAGAGAGAA 3366  
QY 3525 ATGCTTTTAAATGATTCGTACCAACGTTTCTCATGTGTTGGCCAGTTGGATGACATA 3584  
DB 3367 ATGCTTTTAAATGATTCGTACCAACGTTTCTCATGTGTTGGCCAGTTGGATGACATA 3426  
QY 3585 AGAARAACTTAGGAGATTTGTTGCTGAATGACAACTGCAACATTCACCAATCATTAAGAT 3644  
DB 3427 AGAARAACTTAGGAGATTTGTTGCTGAATGACAACTGCAACATTCACCAATCATTAAGAT 3486  
QY 3645 GCTCAGACAGTGAAGGCTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCACTACCTTC 3704  
DB 3487 GCTCAGACAGTGAAGGCTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCACTACCTTC 3546  
QY 3705 CAATTTGAATGCCAAGAGATTCGAAACCGTTTCTCATATGATGATGATGCGAGAA 3764  
DB 3547 CAATTTGAATGCCAAGAGATTCGAAACCGTTTCTCATATGATGATGATGCGAGAA 3606  
QY 3765 TGGAGGCTTATCGAGCAAAATGAAGTTTGGACCCATGCTGCTAGCAACATTCATGAT 3824  
DB 3607 TGGAGGCTTATCGAGCAAAATGAAGTTTGGACCCATGCTGCTAGCAACATTCATGAT 3666  
QY 3825 ATGTTTACTATATCTCATTTTGTCTGACAGTAAATGCACTTAAGCGGAGATATTT 3884  
DB 3667 ATGTTTACTATATCTCATTTTGTCTGACAGTAAATGCACTTAAGCGGAGATATTT 3726  
QY 3885 CCCAGAGGAGGATACACAAGAGCTAGTCCCAATCGATCAGATATAGAGATC 3941  
DB 3727 CCCAGAGGAGGATACACAAGAGCTAGTCCCAATCGATCAGATATAGAGATC 3783

RESULT 12

US-09-895-072-21

; Sequence 21, Application US/09895072

; Patent No. US20020025550A1

GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM M  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASE  
; FILE REFERENCE: 210119USUCONT  
; CURRENT APPLICATION NUMBER: US/09/895,072  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; PRIOR APPLICATION NUMBER: US 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 21  
; LENGTH: 3621  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-09-895-072-21

Query Match 62.3%; Score 3485.4; DB 9; Length 3621;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 3492; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 297 AGCCGAGATCAATACCATGTTTGTGATTCCTATAGAGCAATATGCTGGAAGTCC 356  
DB 115 AGCCGAGATCAATACCATGTTTGTGATTCCTATAGAGCAATATGCTGGAAGTCC 174  
QY 357 TTTCAGAAATCGGCTTTGCTGCCATGCGATTGACGTTGTTTACACCTGGGTGAATGCC 416  
DB 175 TTTCAGAAATCGGCTTTGCTGCCATGCGATTGACGTTGTTTACACCTGGGTGAATGCC 234  
QY 417 ACAGATCTTGAATCTAGGAACTACAGAGGTGACAGAACTAGAGGAGGAGCAG 476  
DB 235 ACAGATCTTGAATCTAGGAACTACAGAGGTGACAGAACTAGAGGAGGAGCAG 294  
QY 477 AAAGCAATGAGAAATCTTGGGAAACACACGGAACCTACTAAGAGAGTGAAG 536  
DB 295 AAAGCAATGAGAAATCTTGGGAAACACACGGAACCTACTAAGAGAGTGAAG 354  
QY 537 CAGTTAGAGTGTTCCTAAACACACTGCAATTAAGGTGCCAATGCTGTACTGACCCAGCC 596  
DB 355 CAGTTAGAGTGTTCCTAAACACACTGCAATTAAGGTGCCAATGCTGTACTGACCCAGCC 414  
QY 597 CTGCGAGCAACATCACCCTGAGGACGTCGATCTCTTTATCTCTTTCTTTCTCTGCTCC 656  
DB 415 CTGCGAGCAACATCACCCTGAGGACGTCGATCTCTTTATCTCTTTCTTTCTTTCTGCT 474  
QY 657 AGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCTCAGTTGTTCT 716  
DB 475 AGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCTCAGTTGTTCT 534  
QY 717 TTGACAGTACTAAGGATGTTGAAGTCCCACTCTGGAAGTCTTAAAGGAAATAGCAGA 776  
DB 535 TTGACAGTACTAAGGATGTTGAAGTCCCACTCTGGAAGTCTTAAAGGAAATAGCAGA 594  
QY 777 CAGACAGTATGGAGGGGTACTTGACACAGATAAAGAGTCCCTGGATAGTGTCTAATG 836  
DB 595 CAGACAGTATGGAGGGGTACTTGACACAGATAAAGAGTCCCTGGATAGTGTCTAATG 654  
QY 837 CAAGATTTGGCTTTCTGAGTGGATTTCCACCAATTCAGAGAAACAAATCAACTAAAA 896  
DB 655 CAAGATTTGGCTTTCTGAGTGGATTTCCACCAATTCAGAGAAACAAATCAACTAAAA 714  
QY 897 ACAGAAATGCGCAAGAAATCTTCTCTAAAGTCAAACTGTTGAGTGTGATTCAGAGGCC 956  
DB 715 ACAGAAATGCGCAAGAAATCTTCTCTAAAGTCAAACTGTTGAGTGTGATTCAGAGGCC 774  
QY 957 AGTGTAGCGCTTCTAAAACTGAATTAACCCCAAGGATTTTCAAGAAATTAAGTAAGCAACT 1016  
DB 775 AGTGTAGCGCTTCTAAAACTGAATTAACCCCAAGGATTTTCAAGAAATTAAGTAAGCAACT 834  
QY 1017 AAGAGAACTGACCATGATGGAAGAACTGACCAATAGTCTGATATTTATTATGG 1076  
DB 835 AAGAGAACTGACCATGATGGAAGAACTGACCAATAGTCTGATATTTATTATGG 894

QY	1136	1137	GATCTGAGCGCCATCAGCCAGTCTTAAGCAGGATGAAGACATCTCTGCCAGTCGTTTGGAA	1138
DB	895	GATCTGAGCGCCATCAGCCAGTCTTAAGCAGGATGAAGACATCTCTGCCAGTCGTTTGGAA	896	954
QY	1137	GATAACGAAGAAGCTGAGGTACTCATTGCGATCTATCGAGAGCATGACCATGGGTGCG	1138	1196
DB	955	GATAACGAAGAAGCTGAGGTACTCATTGCGATCTATCGAGAGCATGACCATGGGTGCG	956	1014
QY	1197	AATATTTTCATTGTCACCAACGGGCAGATTCATCTGGCTGAACTTCGACAACTCCTCGA	1198	1256
DB	1015	AATATTTTCATTGTCACCAACGGGCAGATTCATCTGGCTGAACTTCGACAACTCCTCGA	1016	1074
QY	1257	GTGACAATAGTAACACACAGGATGTTTTTCGAAATTTTGAGCCACTTGCTACCTTTAGT	1258	1316
DB	1075	GTGACAATAGTAACACACAGGATGTTTTTCGAAATTTTGAGCCACTTGCTACCTTTAGT	1076	1134
QY	1317	TCACCTGCTATTGAAAGTCACATTCATCGATCGAAGGCTGTCCAGAAAGTTTATTTC	1318	1376
DB	1135	TCACCTGCTATTGAAAGTCACATTCATCGAAGGCTGTCCAGAAAGTTTATTTC	1136	1194
QY	1377	CTAAATGATGATGTCATGTTTGGGAAGGATGTCGGCCAGATGATTTTACAGTCACTCC	1378	1436
DB	1195	CTAAATGATGATGTCATGTTTGGGAAGGATGTCGGCCAGATGATTTTACAGTCACTCC	1196	1254
QY	1437	AAAGGCCAGAAAGTTTATTGACATGGCTGTGCCAAACTGTGCCAGAGGCTGCCCAGT	1438	1496
DB	1255	AAAGGCCAGAAAGTTTATTGACATGGCTGTGCCAAACTGTGCCAGAGGCTGCCCAGT	1256	1314
QY	1497	TCCTGGATTAAAGATGGCTATTGTGACAAAGCTTGTAATAATTACGCTCGGATTCGGAT	1498	1556
DB	1315	TCCTGGATTAAAGATGGCTATTGTGACAAAGCTTGTAATAATTACGCTCGGATTCGGAT	1316	1374
QY	1557	GGTGGGGAATGCTCTGGAAAACAGTGGAGGAGTCGCTATATTGACAGAGGTGAGGTACT	1558	1616
DB	1375	GGTGGGGAATGCTCTGGAAAACAGTGGAGGAGTCGCTATATTGACAGAGGTGAGGTACT	1376	1434
QY	1617	GGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTCTTAC	1618	1676
DB	1435	GGGAGTATTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTCTTAC	1436	1494
QY	1677	TGTAATCAGGGATGTGCGAATTCCTGGCTCGCTGATAAGTTCTGTGACCAAGCATCGAAT	1678	1736
DB	1495	TGTAATCAGGGATGTGCGAATTCCTGGCTCGCTGATAAGTTCTGTGACCAAGCATCGAAT	1496	1554
QY	1737	GTCTTGTCCTGTGGTTTGTGATGTGGCGACTGTGGGCAAGATCATTTTCATGAATTTGAT	1738	1796
DB	1555	GTCTTGTCCTGTGGTTTGTGATGTGGCGACTGTGGGCAAGATCATTTTCATGAATTTGAT	1556	1614
QY	1797	AAAGTGTATCCTTCTCCAAAACGACATCACTATATTATTCGAAAGGTGAATGCCCTGCCT	1798	1856
DB	1615	AAAGTGTATCCTTCTCCAAAACGACATCACTATATTATTCGAAAGGTGAATGCCCTGCCT	1616	1674
QY	1857	TATTTACAGCTTTGACAGAAAGTAGCCAAAAGAGGAGTTGAAGGTGCTATAGTGACAAATCCA	1858	1916
DB	1675	TATTTACAGCTTTGACAGAAAGTAGCCAAAAGAGGAGTTGAAGGTGCTATAGTGACAAATCCA	1676	1734
QY	1917	ATAATTGCAATGCTTCTATTGCCAACAGTGGNAACCATCCACCTCATTAATGCAACAGT	1918	1976
DB	1735	ATAATTGCAATGCTTCTATTGCCAACAGTGGNAACCATCCACCTCATTAATGCAACAGT	1736	1794
QY	1977	GGAAATGAATGCCACCAACAATACATTTTAAATCTCAAGTTTCAAAATACAAACGATGAAGAG	1978	2036
DB	1795	GGAAATGAATGCCACCAACAATACATTTTAAATCTCAAGTTTCAAAATACAAACGATGAAGAG	1796	1854
QY	2037	TTCAAAATGCAATGAATAACAGTGGAGGTGGACACAGGAGGAGGCCAAACCTGAATTTCTACG	2038	2096
DB	1855	TTCAAAATGCAATGAATAACAGTGGAGGTGGACACAGGAGGAGGCCAAACCTGAATTTCTACG	1856	1914
QY	2097	GCCACAGAGGGTTACGAAAATTTTAGTTAGTCCCTTAACTCTCTCCAGAGGGCGGAATC	2098	2156
DB	1915	GCCACAGAGGGTTACGAAAATTTTAGTTAGTCCCTTAACTCTCTCCAGAGGGCGGAATC	1916	1974

QY	2157	CTTTTGGAGATNTCCCAAAGAAAAACCGTCTCCCGAAGTTTTAAAGACATGATGTTAAC	2216
DB	1975	CTTTTTCAGGATATTCCCRAAGAAAAACCGTCTCCCGAAGTTTTAAAGACATGATGTTAAC	2034
QY	2217	TCAACAAGGAGAGCCACGGAAGAGGTGAAAATTCCTCTGGTAAATATTTCACTCCTTCCA	2276
DB	2035	TCAACAAGGAGAGCCACGGAAGAGGTGAAAATTCCTCTGGTAAATATTTCACTCCTTCCA	2094
QY	2277	AAAGACGCCCCAGTTGAGTCTCAATACTCTTGAGATTTGCAACTGGAACATGGAGACATCACT	2336
DB	2095	AAAGACGCCCCAGTTGAGTCTCAATACTCTTGAGATTTGCAACTGGAACATGGAGACATCACT	2154
QY	2337	TTGAAAGGATACAAATTTGTCCTCAAGTAGCCCTGCTGAGATCATTTCTGATGAATCACAG	2396
DB	2155	TTGAAAGGATACAAATTTGTCCTCAAGTAGCCCTGCTGAGATCATTTCTGATGAATCACAG	2214
QY	2397	CATGCTAAAAATAAAAAATCAAGCTATAATAACAGATGAAGAACAAATGCAGTTTGGTGCT	2456
DB	2215	CATGCTAAAAATAAAAAATCAAGCTATAATAACAGATGAAGAACAAATGCAGTTTGGTGCT	2274
QY	2457	CCACAGAAAAACAGGTTCTATAAAGCATCTTCCCAACAGCTTAGGAGTGTCTGAAAGA	2516
DB	2275	CCACAGAAAAACAGGTTCTATAAAGCATCTTCCCAACAGCTTAGGAGTGTCTGAAAGA	2334
QY	2517	TTGCAGAGGTTGACTTTTCTCTGCAGTGAGTGTAAAAGTGAATGGTCTACACCAGGGTCAG	2576
DB	2335	TTGCAGAGGTTGACTTTTCTCTGCAGTGAGTGTAAAAGTGAATGGTCTACACCAGGGTCAG	2394
QY	2577	AATCCACCCCTGACCTTGGAGACACACAGCAAGATTTAGATGGAATCTCACCCCCAAAA	2636
DB	2395	AATCCACCCCTGACCTTGGAGACACACAGCAAGATTTAGATGGAATCTCACCCCCAAAA	2454
QY	2637	ACCATAGCGCGAATGTGACAAAAGAAAAGCCCCCATCTCTGATTTGTTCACTGGAAAGC	2696
DB	2455	ACCATAGCGCGAATGTGACAAAAGAAAAGCCCCCATCTCTGATTTGTTCACTGGAAAGC	2514
QY	2697	CAGATGCAAAAAAGAAAAAGAAAATCACAGGGAAGAAAAAGAGAACAGTAGATGGAGGAA	2756
DB	2515	CAGATGCAAAAAAGAAAAAGAAAATCACAGGGAAGAAAAAGAGAACAGTAGATGGAGGAA	2574
QY	2757	AATGCTGAAAAATCACATAGCGGTTACTGTAAGTGTACTCTGGAAGAAAGCTGCAGCATTC	2816
DB	2575	AATGCTGAAAAATCACATAGCGGTTACTGTAAGTGTACTCTGGAAGAAAGCTGCAGCATTC	2634
QY	2817	ACAGATAGTTACTTTGGGCTTTTGGCCATGGGAGAAAAAAAAGTATTTCCAAGATCTTCTC	2876
DB	2635	ACAGATAGTTACTTTGGGCTTTTGGCCATGGGAGAAAAAAAAGTATTTCTAGATCTTCTC	2694
QY	2877	GACGAAGAGAGTCATTGGAAGACAAATTTGGCATCTTCTGATACGAAATACTCTGGG	2936
DB	2695	GACGAAGAGAGTCATTGGAAGACAAATTTGGCATCTTCTGATACGAAATACTCTGGG	2754
QY	2937	AGGCAACTAAAGATACATTTGACAGATTCCCTCAGATATGTAAATAAAAATTTCTAAATAGC	2996
DB	2755	AGGCAACTAAAGATACATTTGACAGATTCCCTCAGATATGTAAATAAAAATTTCTAAATAGC	2814
QY	2997	AAGTTTGGATTCACATCGCGAAAGTCCCTGCTCACATGCTCATGATTTGACCGGATT	3056
DB	2815	AAGTTTGGATTCACATCGCGAAAGTCCCTGCTCACATGCTCATGATTTGACCGGATT	2874
QY	3057	GTTATGCAAGAACCTGCAAGATATGTTCCCTGAAGAAATTTGCAAGACGTCATTTTCAAAA	3116
DB	2875	GTTATGCAAGAACCTGCAAGATATGTTCCCTGAAGAAATTTGCAAGACGTCATTTTCAAAA	2934
QY	3117	GTGGGCCATCTGAGGATATGCAGTTTGGCTTCTCTTATTTTATTTATCTCATGAGTGCA	3176
DB	2935	GTGGGCCATCTGAGGATATGCAGTTTGGCTTCTCTTATTTTATTTATCTCATGAGTGCA	2994
QY	3177	GTGAGCCACTGAATATATCTCAAGCTCTTTGATCAAGTTGATACAGATCAATCTGGTGTC	3236
DB	2995	GTGAGCCACTGAATATATCTCAAGCTCTTTGATCAAGTTGATACAGATCAATCTGGTGTC	3054
QY	3237	TTGCTCTGACAGAGAAATCCGAACACTGGCTACCAAGATTCACGAATTCGCGGTTAAGTTTG	3296





1497 TCCTGGATTAAAGGATGGCTATTGTGACAAAGCTTGTAAATAATTCAGCCTGCGATTGGGAT 1556  
1315 TCCTGGATTAAAGGATGGCTATTGTGACAAAGCTTGTAAATAATTCAGCCTGCGATTGGGAT 1374  
1557 GGTGGGATTTCCTCTGGGAAACAGTGGAGGGAGTGGCTATATTTCAGAGAGGTGGAGGTACT 1616  
1375 GGTGGGATTTCCTCTGGGAAACAGTGGAGGGAGTGGCTATATTTCAGAGAGGTGGAGGTACT 1434  
1617 GGGAGTATTGGAGTTGGACACCCCTGGCAGCTTTGGTGGAGGATAAACAAGTGTCTCTTAC 1676  
1435 GGGAGTATTGGAGTTGGACACCCCTGGCAGCTTTGGTGGAGGATAAACAAGTGTCTCTTAC 1494  
1677 TGTAAATCAGGGATGTGGAAATTCCTGGCTCGCTGATTAAGTTCTGTGACCAAGCATGCAAT 1736  
1495 TGTAAATCAGGGATGTGGAAATTCCTGGCTCGCTGATTAAGTTCTGTGACCAAGCATGCAAT 1554  
1737 GTCTGTCTGTGGGTTTGTATGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1796  
1555 GTCTGTCTGTGGGTTTGTATGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1614  
1797 AAAGTGTATCTCTTCCCAAAACCAAGCTCACTATATTATTTCCAAAAGGTGAATGCCTGCT 1856  
1615 AAAGTGTATCTCTTCCCAAAACCAAGCTCACTATATTATTTCCAAAAGGTGAATGCCTGCT 1674  
1857 TATTTAGCTTTGCGAAGTAGCCAAAGAGGAGTTGCAAGTGCCTATAGTGACATCCA 1916  
1675 TATTTAGCTTTGCGAAGTAGCCAAAGAGGAGTTGCAAGTGCCTATAGTGACATCCA 1734  
1917 ATAAATCGACATGCTTCTATTGCGCAACCAAGTGGAAACCACTCACCTCATATAATGCACAGT 1976  
1735 ATAAATCGACATGCTTCTATTGCGCAACCAAGTGGAAACCACTCACCTCATATAATGCACAGT 1794  
1977 GGAATGAATGCCACCAATACATTTTAAATCTCAGCTTTCAAAATACAAACGATGAGAG 2036  
1795 GGAATGAATGCCACCAATACATTTTAAATCTCAGCTTTCAAAATACAAACGATGAGAG 1854  
2037 TTCAAAATGACATTAACAGTGGAGGTGGACCAAGGGAGGACCAAACTGGAATTTAG 2096  
1855 TTCAAAATGACATTAACAGTGGAGGTGGACCAAGGGAGGACCAAACTGGAATTTAG 1914  
2097 GCCCAGAGGGTTACGAAATTTAGTTAGTCCCATACACTTCTCCAGAGGGCGAAATC 2156  
1915 GCCCAGAGGGTTACGAAATTTAGTTAGTCCCATACACTTCTCCAGAGGGCGAAATC 1974  
2157 CTTTTCGAGGATATCCCAAGAAACAGCTTCCCGAAGTTTAAAGACATGATGTTAAC 2216  
1975 CTTTTCGAGGATATCCCAAGAAACAGCTTCCCGAAGTTTAAAGACATGATGTTAAC 2034  
2217 TCACAAAGGAGAGCCCAAGGAGGTGAAATTCCTCGTGGTAAATTTTCACTCTCCCA 2276  
2035 TCACAAAGGAGAGCCCAAGGAGGTGAAATTCCTCGTGGTAAATTTTCACTCTCCCA 2094  
2277 AAAGACGCCAGTTGAGTCTCAATACCTTGGATTGGAACTGGAACTGGAGACATCACT 2336  
2095 AAAGACGCCAGTTGAGTCTCAATACCTTGGATTGGAACTGGAACTGGAGACATCACT 2154  
2337 TTGAAGGATACATATTTGTCCAACTCAGCTTGTGAGATCACTTTCTGATGAATCACTCAG 2396  
2155 TTGAAGGATACATATTTGTCCAACTCAGCTTGTGAGATCACTTTCTGATGAATCACTCAG 2214  
2397 CATGCTAAATTAATAATCAAGCTATTAATAACAGATGAACAAATGACAGTTGGTGGCT 2456  
2215 CATGCTAAATTAATAATCAAGCTATTAATAACAGATGAACAAATGACAGTTGGTGGCT 2274  
2457 CCACAGGAAACAGGTTTCATAAAGCATCTTGCACAAAGCTTAGGAGTGTCTGAAGA 2516  
2275 CCACAGGAAACAGGTTTCATAAAGCATCTTGCACAAAGCTTAGGAGTGTCTGAAGA 2334  
2517 TTGCAGAGGTTGACTTTTCTCGAGTGAAGTGTAAAGTGAATGGTCAATGACAGGCTCAG 2576  
2335 TTGCAGAGGTTGACTTTTCTCGAGTGAAGTGTAAAGTGAATGGTCAATGACAGGCTCAG 2394  
2577 AATCCACCCCTGGACTTGGAGACCACAGCAAGATTTTAGAGTGGAACTCAGACCCCAAAA 2636

2395 AATCCACCCCTGGACTTGGAGACCACAGCAAGATTTTAGAGTGGAACTCAGACCCCAAAA 2454  
2637 ACCATAGCGGAAATGTGACAAAGAAAGCCCCCATCTCTGATTGTTCACCTGGAAAGC 2696  
2455 ACCATAGCGGAAATGTGACAAAGAAAGCCCCCATCTCTGATTGTTCACCTGGAAAGC 2514  
2697 CAGATGACAAAGAAAGAAATACAGAGGAAAGAAAGAAAGACAGTAGAATGGAGGAA 2756  
2515 CAGATGACAAAGAAAGAAATACAGAGGAAAGAAAGAAAGACAGTAGAATGGAGGAA 2574  
2757 AATCTGAAATACATAGAGGTTTACTGAAAGTGTACTTGCAGAAAGCTGACAGCATTAC 2816  
2575 AATCTGAAATACATAGAGGTTTACTGAAAGTGTACTTGCAGAAAGCTGACAGCATTAC 2634  
2817 ACAGATAGTTACTTGGGCTTTTCCCATGGGAGAAAGAAAGATTTTCCAAAGATCTTCTC 2876  
2635 ACAGATAGTTACTTGGGCTTTTCCCATGGGAGAAAGAAAGATTTTCCTAGATCTTCTC 2694  
2877 GACGAGAGAGTCAATTTGAGACACAAATTTGSCATACCTCTCATGATAGCAAAATACTGGG 2936  
2695 GACGAGAGAGTCAATTTGAGACACAAATTTGSCATACCTCTCATGATAGCAAAATACTGGG 2754  
2937 AGGCAACTTAAAGATACATTTTGCAGATTCCTCTCAGATATGTAAATAAAATTTCTAAATPAGC 2996  
2755 AGGCAACTTAAAGATACATTTTGCAGATTCCTCTCAGATATGTAAATAAAATTTCTAAATPAGC 2814  
2997 AAGTTTGGATTACATCGCGGAAAGTCCCTGCTCAGATGCTCAGATGATGACCGGATT 3056  
2815 AAGTTTGGATTACATCGCGGAAAGTCCCTGCTCAGATGCTCAGATGATGACCGGATT 2874  
3057 GTTATGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAGAGAGCTCATTTTCAAAA 3116  
2875 GTTATGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAGAGAGCTCATTTTCAAAA 2934  
3117 GTGCGCCATTCTGAGGATATGAGTGTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3176  
2935 GTGCGCCATTCTGAGGATATGAGTGTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2994  
3177 GTGCGCCATTCTGAGGATATGAGTGTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3236  
2995 GTGCGCCATTCTGAGGATATGAGTGTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3054  
3237 TTGTCTGACAGAGAAATCCGAACTGCTACCAAGATTCACGAACTGCGGTTAAGTTTG 3296  
3055 TTGTCTGACAGAGAAATCCGAACTGCTACCAAGATTCACGAACTGCGGTTAAGTTTG 3114  
3297 CAGGTTTGAAGGCTTGGAGACATGCTTAATTAATTTGCTCAAAATGCTTCTCTGCTGAT 3356  
3115 CAGGTTTGAAGGCTTGGAGACATGCTTAATTAATTTGCTCAAAATGCTTCTCTGCTGAT 3174  
3357 ATCAGCAGCTTAAATTAATTTCCCACTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3416  
3175 ATCAGCAGCTTAAATTAATTTCCCACTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3234  
3417 CCGGTCATTAAGTCTAGTAACTGTAACCTGTAACCTGTAACCTGTAACCTGTAACCTGTAAC 3476  
3235 CCGGTCATTAAGTCTAGTAACTGTAACCTGTAACCTGTAACCTGTAACCTGTAACCTGTAAC 3294  
3477 TATAAGGACAAAGAAATAGGTTTGAATCATGGAAGAAAGAAATGCGCTTTTAAA 3536  
3295 TATAAGGACAAAGAAATAGGTTTGAATCATGGAAGAAAGAAATGCGCTTTTAAA 3354  
3537 ATGATTCTGTAACAGGTTTCTCATGTGGTGGCAGTGGATGATGAAGAAAGAAAGCT 3596  
3355 ATGATTCTGTAACAGGTTTCTCATGTGGTGGCAGTGGATGATGAAGAAAGAAAGCT 3414  
3597 AGGAGTTTGTGCTGTAATGACAACTGACCAATGACCAATGACCAATGACCAATGACCAATGAC 3656  
3415 AGGAGTTTGTGCTGTAATGACAACTGACCAATGACCAATGACCAATGACCAATGACCAATGAC 3474  
3657 AAGGCTGTCTCAGGAGCTTCTATGAATCCATGTTCCTCCATGCTTCCCAATTTTGAAGCTG 3716



Db 3475 AAGCTGTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAAGT 3534  
QY 3717 CCAAGAGATGCAAAACGTTTCTTCATATGATGATGAGTGGAGGAGTGGAGGCTTAT 3776  
Db 3535 CCAAGAGATGCAAAACGTTTCTTCATATGATGATGAGTGGAGGAGTGGAGGCTTAT 3594  
QY 3777 CGAGACAAATGAAGTTTGTGAC 3799  
Db 3595 CGAGACAAATGAAGTAGTAGTC 3617

RESULT 14  
US-10-306-686-21  
; Sequence 21, Application US/10306686  
; Publication No. US20030148460A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM  
; TITLE OF INVENTION: PHOSPHODIESTER ALPHA-GLUCNACASE OF THE LYSOSOMAL TARGETING PATHWAY  
; FILE REFERENCE: 230397US77DIV  
; CURRENT APPLICATION NUMBER: US/10/306,686  
; CURRENT FILING DATE: 2002-11-29  
; PRIOR APPLICATION NUMBER: 09/636,596  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-08-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 21  
; LENGTH: 3621  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-306-686-21

Query Match 62.3%; Score 3485.4; DB 15; Length 3621;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 3492; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 297 AGCGAGATCAATACCATGTTTCTGATTCCTATAGACATATATGCTGGAAAGTCC 356  
Db 115 AGCGAGATCAATACCATGTTTCTGATTCCTATAGACATATATGCTGGAAAGTCC 174  
QY 357 TTTGAGATCGGCTTGTCTGCCCATGCCATGCGGATGCGGTTGTTTACACCTGGGTGAATGGC 416  
Db 175 TTTGAGATCGGCTTGTCTGCCCATGCCATGCGGATGCGGTTGTTTACACCTGGGTGAATGGC 234  
QY 417 ACAGATCTTGAACTACTGAGGAATACAGCAGGTGAGAACAGATGAGGAGGAGCAG 476  
Db 235 ACAGATCTTGAACTACTGAGGAATACAGCAGGTGAGAACAGATGAGGAGGAGCAG 294  
QY 477 AAAGCAATGAGAGAAATCTTGGGAAACACACACGAACTCTACTAAGAAAGAGTGAGAAG 536  
Db 295 AAAGCAATGAGAGAAATCTTGGGAAACACACACGAACTCTACTAAGAAAGAGTGAGAAG 354  
QY 537 CAGTTAGAGTGTGCTAAACACATGCAATTAAGGTGCAATGCTTGACTGGACCCAGCC 596  
Db 355 CAGTTAGAGTGTGCTAAACACATGCAATTAAGGTGCAATGCTTGACTGGACCCAGCC 414  
QY 597 CTGCGCCCAATCACCTGAGGAGGTGCGCATCTCTTATCTCTTCTCTTCTCTCTCTCTCTCT 656  
Db 415 CTGCGCCCAATCACCTGAGGAGGTGCGCATCTCTTATCTCTTCTCTCTCTCTCTCTCTCT 474  
QY 657 AGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCTCAGTTGTTGTT 716  
Db 475 AGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCTCAGTTGTTGTT 534  
QY 717 TTTGACAGTACTAGGATGTTGAAGATGCCCACTCTGAGCTGCTTAAAGAAATAGCAGA 776  
Db 535 TTTGACAGTACTAGGATGTTGAAGATGCCCACTCTGAGCTGCTTAAAGAAATAGCAGA 594  
QY 777 CAGACAGTATGAGGGGTACTTGCACACAGATTAAGAGTCCCTGATGATGCTTAATG 836  
Db 595 CAGACAGTATGAGGGGTACTTGCACACAGATTAAGAGTCCCTGATGATGCTTAATG 654

QY 837 CAAGATTTGGCTTCTCTGAGTGGATTTCCACCAACATTCAGAGAAACAAATCAACTAAAA 896  
Db 655 CAAGATTTGGCTTCTCTGAGTGGATTTCCACCAACATTCAGAGAAACAAATCAACTAAAA 714  
QY 897 ACAAATTTCCAGAAATCTTCTCTAAAGTCAAACTGTTGCACTGTTGATTCAGAGGCC 956  
Db 715 ACAAATTTCCAGAAATCTTCTCTAAAGTCAAACTGTTGCACTGTTGATTCAGAGGCC 774  
QY 957 AGTGTAGCGCTTCTAAACCTGAATAACCCCAAGGATTTTCAAGAAATTAAGAAACAACT 1016  
Db 775 AGTGTAGCGCTTCTAAACCTGAATAACCCCAAGGATTTTCAAGAAATTAAGAAACAACT 834  
QY 1017 AAGAAGACATGACCATTTGATGGAAGAAAGTGAACATTAAGTCTGCAATTTATATGG 1076  
Db 835 AAGAAGACATGACCATTTGATGGAAGAAAGTGAACATTAAGTCTGCAATTTATATGG 894  
QY 1077 GATCTGAGCGCCATCAGCCAGTCTAAGCAGATGAAGCATCTCTGCCAGTGTGTTTGA 1136  
Db 895 GATCTGAGCGCCATCAGCCAGTCTAAGCAGATGAAGCATCTCTGCCAGTGTGTTTGA 954  
QY 1137 GATAACGAAAGTCTGAGTACTCATTTGCGATCTATCGAGGAGTGCACCATGGTTCGG 1196  
Db 955 GATAACGAAAGTCTGAGTACTCATTTGCGATCTATCGAGGAGTGCACCATGGTTCGG 1014  
QY 1197 AATATTTTCAATGTCACCAACCGGAGATTCCTCTGCTGAACCTTGACATCTCTGA 1256  
Db 1015 AATATTTTCAATGTCACCAACCGGAGATTCCTCTGCTGAACCTTGACATCTCTGA 1074  
QY 1257 GTGACAAATAGTAACACACAGGATGTTTTTGAATTTTGAAGCACTTGCTACCTTTAGT 1316  
Db 1075 GTGACAAATAGTAACACACAGGATGTTTTTGAATTTTGAAGCACTTGCTACCTTTAGT 1134  
QY 1317 TCACCTGCTATTGAAAGTCAATTCATGTCATCGAAGGCTGCTCCAGAAAGTTTATTTAC 1376  
Db 1135 TCACCTGCTATTGAAAGTCAATTCATGTCATCGAAGGCTGCTCCAGAAAGTTTATTTAC 1194  
QY 1377 CTAATATGATGATGTCATGTTTGGGAGGATGCTGCCAGATGATTTTACAGTCACTCC 1436  
Db 1195 CTAATATGATGATGTCATGTTTGGGAGGATGCTGCCAGATGATTTTACAGTCACTCC 1254  
QY 1437 AAAGCCAGAAAGTTTATTTGACATGCGCTGTGCCAAACTGTGTGCCAGGCTGCCAGGT 1496  
Db 1255 AAAGCCAGAAAGTTTATTTGACATGCGCTGTGCCAAACTGTGTGCCAGGCTGCCAGGT 1314  
QY 1497 TCCTGATTAAGAGTGGCTATTGTCAGAGGCTGCTTAATTAATTCAGCTCGGATGGGAT 1556  
Db 1315 TCCTGATTAAGAGTGGCTATTGTCAGAGGCTGCTTAATTAATTCAGCTCGGATGGGAT 1374  
QY 1557 GGTGGGATTTGCTCTGGAACACAGTGGAGGAGTGCCTATATTTCAGGAGGTGGAGTACT 1616  
Db 1375 GGTGGGATTTGCTCTGGAACACAGTGGAGGAGTGCCTATATTTCAGGAGGTGGAGTACT 1434  
QY 1617 GGGAGTATGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGATTAACAGTGTCTCTTAC 1676  
Db 1435 GGGAGTATGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGATTAACAGTGTCTCTTAC 1494  
QY 1677 TGTAACTCAGGATGTGCGAATTTCTTGGCTGCTGATTAAGTTCTGTGACCAAGCATGCAAT 1736  
Db 1495 TGTAACTCAGGATGTGCGAATTTCTTGGCTGCTGATTAAGTTCTGTGACCAAGCATGCAAT 1554  
QY 1737 GTCTGTCTGTGGTGTGATGCTGGGAGTGTGGGAGATCATTTTTCATGATTTGAT 1796  
Db 1555 GTCTGTCTGTGGTGTGATGCTGGGAGTGTGGGAGATCATTTTTCATGATTTGAT 1614  
QY 1797 AAGTGTATCTTCTCCCAACAGACTCACTATATTATTTCCAAAAGGTGAATGCTGCT 1856  
Db 1615 AAGTGTATCTTCTCCCAACAGACTCACTATATTATTTCCAAAAGGTGAATGCTGCT 1674  
QY 1857 TATTTTCACTTTGCAAGAGTACCCAAAGAGAGGAGTGAAGGTGCTCTATAGTACAAAT 1916  
Db 1675 TATTTTCACTTTGCAAGAGTACCCAAAGAGAGGAGTGAAGGTGCTCTATAGTACAAAT 1734  
QY 1917 ATAATTCGACATGCTTCTATTTCGCAACAGTGGAAACCAATCCACTCTATATGACAGT 1976



Query Match				61.9%; Score 3465.6; DB 15; Length 3600;			
Best Local Similarity				99.5%; Pred. No. 0;			
Matches 3477; Conservative				0; Mismatches 19; Indels 0; Gaps 0;			
QY	297	AGCGAGATCAATACCATGTTTGTGTTGATTCCTATAGACAAATATGCTGGAAGTCC	356				
Db	103	AGCGAGATCAATACCATGTTTGTGTTGATTCCTATAGACAAATATGCTGGAAGTCC	162				
QY	357	TTTCAGAAATCGGCTTTGTCTCCCATGCGGATTCGCGATTCGTTTACACTCGGTGAAATGGC	416				
Db	163	TTTCAGAAATCGGCTTTGTCTCCCATGCGGATTCGCGATTCGTTTACACTCGGTGAAATGGC	222				
QY	417	ACAGATCTTGAATCTAGTGAAGAACTACAGCAGTTCAGAGAAACAGATGGAAGGAGCAG	476				
Db	223	ACAGATCTTGAATCTAGTGAAGAACTACAGCAGTTCAGAGAAACAGATGGAAGGAGCAG	282				
QY	477	AAAGCAATAGAGAAATCTTTGGGAAACACAAACGGAACCTACTAAGAAAGATGAGAAG	536				
Db	283	AAAGCAATAGAGAAATCTTTGGGAAACACAAACGGAACCTACTAAGAAAGATGAGAAG	342				
QY	537	CAGTTAGAGTCTTGTCTAACACATGCAATTAAGTGGCCATGCTTGTACTGGACCCAGCC	596				
Db	343	CAGTTAGAGTCTTGTCTAACACATGCAATTAAGTGGCCATGCTTGTACTGGACCCAGCC	402				
QY	597	CTGCCAGCAACATCACCTGAAGGACGTGCCATCTCTTTATCTCTCTTTTCAATCTTGCC	656				
Db	403	CTGCCAGCAACATCACCTGAAGGACGTGCCATCTCTTTATCTCTCTTTTCAATCTTGCC	462				
QY	657	AGTGACATTTTCAATGTTGGAACACCAAAACCCCTTCTACCAATGCTCAGTTGTTGT	716				
Db	463	AGTGACATTTTCAATGTTGGAACACCAAAACCCCTTCTACCAATGCTCAGTTGTTGT	522				
QY	717	TTTGACAGTACTAAGAGTGTGAAGATGCCCACTCTGACCTGCTTAAAGGAAATAGCAGA	776				
Db	523	TTTGACAGTACTAAGAGTGTGAAGATGCCCACTCTGACCTGCTTAAAGGAAATAGCAGA	582				
QY	777	CAGACAGTATGGAAGGGTACTTGACAAACAGATTAAGAGTCCCTGGATAGTCTAATG	836				
Db	583	CAGACAGTATGGAAGGGTACTTGACAAACAGATTAAGAGTCCCTGGATAGTCTAATG	642				
QY	837	CAAGATTTGGCTTCTCTGAGTGGATTTCCACCAACATTCAGAGGAAACAAATCAACTAAA	896				
Db	643	CAAGATTTGGCTTCTCTGAGTGGATTTCCACCAACATTCAGAGGAAACAAATCAACTAAA	702				
QY	897	ACAAATTTGCCAGAAATCTTTCTCTAAAGTCAAACTGTTGTCAGTGTGATTCAGAGGCC	956				
Db	703	ACAAATTTGCCAGAAATCTTTCTCTAAAGTCAAACTGTTGTCAGTGTGATTCAGAGGCC	762				
QY	957	AGTGAGCGCTTTAAACTGAATAAATCCCAAGATTTTCAAGATTTGAATTAAGCAACT	1016				
Db	763	AGTGAGCGCTTTCTAAATCTGAATAAATCCCAAGATTTTCAAGATTTGAATTAAGCAACT	822				
QY	1017	AAGAAGAACATGACCATTTGATGGAAGAACTGACCAATAGTCTGTCATATTTATATGG	1076				
Db	823	AAGAAGAACATGACCATTTGATGGAAGAACTGACCAATAGTCTGTCATATTTATATGG	882				
QY	1077	GATCTGAGCGGCATCAGCAGTCTAAGCAGATGAAGACATCTCTGCCAGTCTGTTTTGAA	1136				
Db	883	GATCTGAGCGGCATCAGCAGTCTAAGCAGATGAAGACATCTCTGCCAGTCTGTTTTGAA	942				
QY	1137	GATACGAGAACTGAGTACTCATGCGATCTATCGAGGGCATGACCATGTTGTTGCG	1196				
Db	943	GATACGAGAACTGAGTACTCATGCGATCTATCGAGGGCATGACCATGTTGTTGCG	1002				
QY	1197	AATATTTTCAATTTGTCACCAAGGGCAGATTCATCTGCTGTAACCTTTGACAATCTCGA	1256				
Db	1003	AATATTTTCAATTTGTCACCAAGGGCAGATTCATCTGCTGTAACCTTTGACAATCTCGA	1062				
QY	1257	GTGACAAATAGTAACACACAGATGTTTTTTCGAATTTGAGCAGCTTGCCTACCTTAGT	1316				
Db	1063	GTGACAAATAGTAACACACAGATGTTTTTTCGAATTTGAGCAGCTTGCCTACCTTAGT	1122				
QY	1317	TCACCTGTATTTGAAAGTCACATTCATCGCATCGAAGGGCTGTCCGAGAAGTTTATTTAC	1376				

Db	1123	TCACCTGTATTTGAAAGTCACATTCATCGCATCGAAGGGCTGTCCGAGAAGTTTATTTAC	1182				
QY	1377	CTAAATGATGATGTATGTTTGGGAAGGATGTCTGGCCAGATGATTTTTCAGTCACTCC	1436				
Db	1183	CTAAATGATGATGTATGTTTGGGAAGGATGTCTGGCCAGATGATTTTTCAGTCACTCC	1242				
QY	1437	AAAGCCAGAAAGTTTATTTGACATGCTGTGCCAATCTGTGCCAGGGCTGCCAGGT	1496				
Db	1243	AAAGCCAGAAAGTTTATTTGACATGCTGTGCCAATCTGTGCCAGGGCTGCCAGGT	1302				
QY	1497	TCTCGATTAAGGATGGCTATTTGTGACAAAGGCTTCTTAATAATTCAGCCTGCGATTTGGAT	1556				
Db	1303	TCTCGATTAAGGATGGCTATTTGTGACAAAGGCTTCTTAATAATTCAGCCTGCGATTTGGAT	1362				
QY	1557	GGTGGGATGCTCTGGGAAACAGTGGAGGAGTCTCTATATTTGCGAGGAGTGGAGTACT	1616				
Db	1363	GGTGGGATGCTCTGGGAAACAGTGGAGGAGTCTCTATATTTGCGAGGAGTGGAGTACT	1422				
QY	1617	GGGAGTATTTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAAATAAACAGTGTCTCTTAC	1676				
Db	1423	GGGAGTATTTGGAGTTGGACACCCCTGGCAGTTTGGTGGAGGAAATAAACAGTGTCTCTTAC	1482				
QY	1677	TGTAATCAGGAGTGTGCGAATTCCTGGCTGCTGTATAGTTCTGTGACCAAGCATGCAAT	1736				
Db	1483	TGTAATCAGGAGTGTGCGAATTCCTGGCTGCTGTATAGTTCTGTGACCAAGCATGCAAT	1542				
QY	1737	GTCTGTGCTGTGGGTTTGTATCTGCTGGCAGTCTGGGCAAGATCATTTTCATGAATGTAT	1796				
Db	1543	GTCTGTGCTGTGGGTTTGTATCTGCTGGCAGTCTGGGCAAGATCATTTTCATGAATGTAT	1602				
QY	1797	AAAGTGATCTCTCTCCAAACAGACTCACTATATTTATTTCCAAAAGGTGAATTCCTGCT	1856				
Db	1603	AAAGTGATCTCTCTCCAAACAGACTCACTATATTTATTTCCAAAAGGTGAATTCCTGCT	1662				
QY	1857	TATTTAGCTTTTGCAGAACTAGCCAAAGAGAGTGAAGTGCCTATAGTCAGCAATCCA	1916				
Db	1663	TATTTAGCTTTTGCAGAACTAGCCAAAGAGAGTGAAGTGCCTATAGTCAGCAATCCA	1722				
QY	1917	ATAAATTCGACATGCTCTTATTTGCCAACAGTGGAAAAACCATCCACTCATATATGCACAGT	1976				
Db	1723	ATAAATTCGACATGCTCTTATTTGCCAACAGTGGAAAAACCATCCACTCATATATGCACAGT	1782				
QY	1977	GGATGAATGCCACCAATACATTTTAAATCTCAGCTTTCAAAATACAAACGATGAAGAG	2036				
Db	1783	GGATGAATGCCACCAATACATTTTAAATCTCAGCTTTCAAAATACAAACGATGAAGAG	1842				
QY	2037	TTCAAAATCAGATTAACAGTGGAGTGGACAAAGGAGGACCAAACTGAATTTCTAGC	2096				
Db	1843	TTCAAAATCAGATTAACAGTGGAGTGGACAAAGGAGGACCAAACTGAATTTCTAGC	1902				
QY	2097	GCCAGAGGGTTTACGAAATTTAGTTAGTCTCCATAACACTTCTTCCAGAGCGGAAATC	2156				
Db	1903	GCCAGAGGGTTTACGAAATTTAGTTAGTCTCCATAACACTTCTTCCAGAGCGGAAATC	1962				
QY	2157	CTTTTTCAGGATATTTCCGAAAGAAACGCTTCCGAGTTTAAAGAGACATGATTAAC	2216				
Db	1963	CTTTTTCAGGATATTTCCGAAAGAAACGCTTCCGAGTTTAAAGAGACATGATTAAC	2022				
QY	2217	TCAAACAGAGAGAGCCAGGAAGAGTGAATAATTCCTCGTGAATAATTTTCACTCTCTCA	2276				
Db	2023	TCAAACAGAGAGAGCCAGGAAGAGTGAATAATTCCTCGTGAATAATTTTCACTCTCTCA	2082				
QY	2277	AAAGAGCCGAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAACTGGAGACATCACT	2336				
Db	2083	AAAGAGCCGAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAACTGGAGACATCACT	2142				
QY	2337	TTGAAAGGATACAAATTTGTCAGAGTCAAGCTTGTGAGATCATTTCTGTGAATCAACAG	2396				
Db	2143	TTGAAAGGATACAAATTTGTCAGAGTCAAGCTTGTGAGATCATTTCTGTGAATCAACAG	2202				
QY	2397	CATGCTTAAATAAAAATCAAGCTATTAATAACAGATGAAACAAATGACAGTTTGTGCT	2456				

Db 2203 CATGCTAAATAAAAAATCAAGCTATATAACAGATGAACAAATGACAGCTTTGGTGGCT 2262  
Qy 2457 CCACAGGAAACAGGTTTCATAAAGCATCTTGCCAAACAGCTTAGGAGTGTCTGAAAGA 2516  
Db 2263 CCACAGGAAACAGGTTTCATAAAGCATCTTGCCAAACAGCTTAGGAGTGTCTGAAAGA 2322  
Qy 2517 TTGCAGAGGTTGACTTTTCTCGAGTGAAGTGAATGGTCAATGACCAAGGCTCAG 2576  
Db 2323 TTGCAGAGGTTGACTTTTCTCGAGTGAAGTGAATGGTCAATGACCAAGGCTCAG 2382  
Qy 2577 AATCCACCCCTGGACTTGGAGACCACAGCAAGATTTAGAGTGAAGTCAACCCAAAAA 2636  
Db 2383 AATCCACCCCTGGACTTGGAGACCACAGCAAGATTTAGAGTGAAGTCAACCCAAAAA 2442  
Qy 2637 ACCATAGCGGAATGTGACAAAAGAAAGCCCCATCTCTGATTTGCCACTGGAAAGC 2696  
Db 2443 ACCATAGCGGAATGTGACAAAAGAAAGCCCCATCTCTGATTTGCCACTGGAAAGC 2502  
Qy 2697 CAGATGACAAAGAAAGAAATCACAGGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 2756  
Db 2503 CAGATGACAAAGAAAGAAATCACAGGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 2562  
Qy 2757 AATGCTGAATACATAGCGGTTACTGAAGTGTACTTGGAGAGAGCTGACGATTAAC 2816  
Db 2563 AATGCTGAATACATAGCGGTTACTGAAGTGTACTTGGAGAGAGCTGACGATTAAC 2622  
Qy 2817 ACAGATAGTTACTTTGGGCTTTTGGCCATGGGAGAAAAAAGATTTTCCAAAGATCTTCTC 2876  
Db 2623 ACAGATAGTTACTTTGGGCTTTTGGCCATGGGAGAAAAAAGATTTTCTAGATCTTCTC 2682  
Qy 2877 GACGAGAGAGTCAATTTGAGACACAAATTTGGCATCTTCACTGATAGCAAAAATCTGGG 2936  
Db 2683 GACGAGAGAGTCAATTTGAGACACAAATTTGGCATCTTCACTGATAGCAAAAATCTGGG 2742  
Qy 2937 AGGCACTAAAGAGATACATTTGCAGATTCCTCAGATATGTAATAAATTTCTAAATAGC 2996  
Db 2743 AGATACAGAGAGATACATTTGCAGATTCCTCAGATATGTAATAAATTTCTAAATAGC 2802  
Qy 2997 AAGTTTGGATTACATCCGGGAAAGTCCCTGCTCACATGCTCACATGATGACCGGATT 3056  
Db 2803 AAGTTTGGATTACATCCGGGAAAGTCCCTGCTCACATGCTCACATGATGACCGGATT 2862  
Qy 3057 GTTATGCAAGAACTGCAAGATATGTTCCCTGAGAAATTTGACAGAGTCAATTTCAAAA 3116  
Db 2863 GTTATGCAAGAACTGCAAGATATGTTCCCTGAGAAATTTGACAGAGTCAATTTCAAAA 2922  
Qy 3117 GTGGCCCATCTGAGGATATGCAAGTTTGCCTTCTCTTATTTATTTATCTCATGAGTGA 3176  
Db 2923 GTGGCCCATCTGAGGATATGCAAGTTTGCCTTCTCTTATTTATTTATCTCATGAGTGA 2982  
Qy 3177 GTGCAGCCACTGATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTGC 3236  
Db 2983 GTGCAGCCACTGATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTGC 3042  
Qy 3237 TTGTCTGACAGAGAAATCCGAAACACTGGCTACAGAAATTCAGAACTGCCGTTAAGTTTG 3296  
Db 3043 TTGTCTGACAGAGAAATCCGAAACACTGGCTACAGAAATTCAGAACTGCCGTTAAGTTTG 3102  
Qy 3297 CAGGATTTGACAGGTTCTGGAAACACATGCTAATAAATTTGCTCAAAAATGCTTCTGCTGAT 3356  
Db 3103 CAGGATTTGACAGGTTCTGGAAACACATGCTAATAAATTTGCTCAAAAATGCTTCTGCTGAT 3162  
Qy 3357 ATCACGAGCTAAATAATATTTCCACCAACTCAGGAATCTCTATGATCCCAACCTGCCA 3416  
Db 3163 ATCACGAGCTAAATAATATTTCCACCAACTCAGGAATCTCTATGATCCCAACCTGCCA 3222  
Qy 3417 CCGGTCACTAAAAGTCTAGTAACAAACTGTAAACCAAGTAACTGACAAAATCCACAAAGCA 3476  
Db 3223 CCGGTCACTAAAAGTCTAGTAACAAACTGTAAACCAAGTAACTGACAAAATCCACAAAGCA 3282  
Qy 3477 TATAAGGACAAAACAAATATAGGTTTGAATCAATGGGAGAGAAATCGCTTTTAAA 3536  
Db 3283 TATAAGGACAAAACAAATATAGGTTTGAATCAATGGGAGAGAAATCGCTTTTAAA 3342

Qy 3537 ATGATTCTGACCAAGTTTCTCATGTGTTGGCCAGTTGGATGACATAAGAAAAACCCCT 3596  
Db 3343 ATGATTCTGACCAAGTTTCTCATGTGTTGGCCAGTTGGATGACATAAGAAAAACCCCT 3402  
Qy 3597 AGGAAAGTTTGTGCTGCTGAATGACAAACATTTGACCAACATCAAAAGATGCTCAGACAGTG 3656  
Db 3403 AGGAAAGTTTGTGCTGCTGAATGACAAACATTTGACCAACATCAAAAGATGCTCAGACAGTG 3462  
Qy 3657 AAGGCTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAACTG 3716  
Db 3463 AAGGCTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAACTG 3522  
Qy 3717 CCAAGAGAGTATCGAAACCGTTTCTTTCATATGATGAGCTGAGGAAATGGAGGGCTTAT 3776  
Db 3523 CCAAGAGAGTATCGAAACCGTTTCTTTCATATGATGAGCTGAGGAAATGGAGGGCTTAT 3582  
Qy 3777 CGAGCAAAATTTGAAGT 3792  
Db 3583 CGAGCAAAATTTGAAGT 3598

Search completed: November 22, 2004, 16:51:27  
Job time : 2633 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2004, 22:56:56 ; Search time 15702 Seconds  
(without alignments)  
12988.994 Million cell updates/sec

Title: US-10-023-888-3  
Perfect score: 5597  
Sequence: 1 cggagcgcagggcgctcg.....aaaagttaattttgaaa 5597

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_hcc.\*  
4: gb\_est3.\*  
5: gb\_est4.\*  
6: gb\_est5.\*  
7: gb\_est6.\*  
8: gb\_gsl1.\*  
9: gb\_gsl2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	856.2	15.3	934	5	BQ719691 AGENCOURT
2	829.6	14.8	1016	4	BM544697 AGENCOURT
3	813	14.5	994	4	BM476404 AGENCOURT
4	812	14.5	981	5	BQ053143 AGENCOURT
5	804.8	14.4	1118	4	BM452210 AGENCOURT
6	794.8	14.2	855	6	CD657400 AGENCOURT
7	771.8	13.8	1061	4	BM450679 AGENCOURT
8	770	13.8	854	5	BUI17138 AGENCOURT
9	751.8	13.4	867	6	CD657425 AGENCOURT
10	746.8	13.3	929	5	BQ720582 AGENCOURT
11	743.8	13.3	783	4	BQ214021 AGENCOURT
12	733.4	13.1	819	6	CD652380 AGENCOURT
13	728.4	13.0	826	5	BUI933148 AGENCOURT
14	715.2	12.8	773	4	BG196325 AGENCOURT
15	710.8	12.7	795	4	BG194321 AGENCOURT
16	694.2	12.4	899	6	CD244899 AGENCOURT
17	692	12.4	780	1	AUI134813 AGENCOURT
18	680.4	12.2	814	6	CD657278 AGENCOURT
19	671	12.0	984	4	BG250072 AGENCOURT
20	665.2	11.9	778	6	CD643692 AGENCOURT
21	664.8	11.9	705	5	BX095637 AGENCOURT
22	644.8	11.5	686	5	BQ008614 UI-H-ED1-
23	643.8	11.5	792	4	BG200937 RST20259
24	637.6	11.4	682	6	CA438470 UI-H-DT1-

25	636.8	11.4	794	6	CD656815
26	623.8	11.1	852	4	BG199915
27	622.6	11.1	630	7	CN426009 170006000
28	620.6	11.1	842	4	BG207173 RST26640
29	617.8	11.0	773	7	CF406980 CH3-M-045 F
30	617.2	11.0	634	5	EX645182 DKF2D781D
31	608	10.9	670	1	AUI135455
32	605.6	10.8	789	7	CN525214 UI-M-GHO-
33	600	10.7	783	7	CK781711 UI-M-GHO-
34	599.8	10.7	613	4	BG163659 602338839
35	599.4	10.7	791	6	CB518925 UI-M-GHO-
36	598.4	10.7	639	7	CK002110 AGENCOURT
37	596	10.6	793	6	CB27476 UI-M-FY0-
38	593.2	10.6	776	7	CF406981 CH3H045 F
39	589.2	10.5	616	4	BM790504 K-EST0070
40	588.6	10.5	819	6	CA324819 UI-M-FY0-
41	584.6	10.4	731	5	BU709362 UI-M-EVO-
42	575	10.3	663	7	CO261199 413319 B
43	573.2	10.2	612	4	BG577274 602595465
44	573	10.2	782	5	BU451401 603215493
45	568.8	10.2	1037	5	BQ226717 AGENCOURT

ALIGNMENTS

RESULT 1  
BQ719691  
LOCUS BQ719691 934 bp mRNA linear EST 16-JUL-2002  
DEFINITION AGENCOURT\_8304671 Lupski\_sympathetic\_trunk Homo sapiens CDNA clone  
IMAGE:6193737 5', mRNA sequence.  
ACCESSION BQ719691  
VERSION BQ719691  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 934)  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
JOURNAL NIH-MGC http://mgc.nci.nih.gov/.  
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: Dr. James R. Lupski  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM13597 row: n column: 10  
High quality sequence stop: 626.  
Location/Qualifiers  
1. .934  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6193737"  
/sex="male"  
/tissue\_type="sympathetic trunk"  
/dev\_stage="adult, 16 yr"  
/lab\_host="DH10B"  
/clone\_lib="Lupski\_sympathetic trunk"  
/note="Vector: pCMV-SPORT6 (Life Technologies); Site\_1:  
NotI; Site\_2: SalI; cDNA made by oligo-dT priming.  
Directionally cloned using the following adaptors:  
5'-TCGACCCACGCTCCG-3' and  
5'-GACTAGTTCAGTCGCGAGCGCCGCTT(15)-3'. Size selected >  
1 kb for average insert length 1.9 kb. This is a primary  
library, non-amplified. Library constructed by Life  
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor  
College of Medicine); available through Life

FEATURES  
source

BM544697	BM544697	1016 bp	mRNA	linear	EST 20-FEB-2002
LOCUS	AGENCOURT_6494522 NIH_MGC_124	Homo sapiens	cdna	clone	IMAGE:5727411
DEFINITION	5', mRNA sequence.				
ACCESSION	BM544697				
VERSION	BM544697.1	GI:18776157			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	NIH-MGC <a href="http://mgs.nci.nih.gov/">http://mgs.nci.nih.gov/</a> .				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999).				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgabbs-remail.nih.gov">cgabbs-remail.nih.gov</a> Tissue Procurement: Invitrogen CDNA Library Preparation: Life Technologies, Inc. DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LLAM12721 row: h column: 10 High quality sequence stop: 690.				
FEATURES	Location/Qualifiers				
source	1..1016				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:5727417"				
	/tissue_type="hippocampus"				
	/lab_host="DH10B"				
	/clone_lib="NIH MGC 124"				
	/notes="Organ: brain; Vector: pCMV-SPORT6; Site 1: EcoRV (destroyed); Site 2: NotI; RNA source male hippocampus, age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 0.9-4 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 012."				
ORIGIN					
Query Match	14.8%;	Score 829.6;	DB 4;	Length 1016;	
Best Local Similarity	97.6%;	Pred. No. 6e-163;			
Matches	863;	Conservative	0;	Mismatches 19;	Indels 2; Gaps 2;
QY	193	CCTATACCTGCTGCCACAGGTATGGCTCTACGTGTGCTTCTGGCGCTGTTGTGCA	252		
Db	1	CCTATACCTGCTGCCACAGGTATGGCTCTACGTGTGCTTCTGGCGCTGTTGTGCA	60		
QY	253	CCATCGTCTCCGCTTCCAGTTCGAGAGTGTTCTGGAATGGAGCCGAGATCAATACC	312		
Db	61	CCATCGTCTCCGCTTCCAGTTCGAGAGTGTTCTGGAATGGAGCCGAGATCAATACC	120		
QY	313	ATGTTTTGTTGATTCTCTATAGAGACAATATTGCTGGAAGTCTTTTCAGAAATCGGCTTT	372		
Db	121	ATGTTTTGTTGATTCTCTATAGAGACAATATTGCTGGAAGTCTTTTCAGAAATCGGCTTT	180		
QY	373	GTCTGCCATCCGATTCAGTTGTTTACACCTGGGTGAATGGCAGAGATCTTGAACCTAC	432		
Db	181	GTCTGCCATCCGATTCAGTTGTTTACACCTGGGTGAATGGCAGAGATCTTGAACCTAC	240		
QY	433	TGAGGAACTACAGCAGGTCAGAGAACAGATGGAGGAGGAGCAGAAAGCAATGAGAGAAA	492		
Db	241	TGAGGAACTACAGCAGGTCAGAGAACAGATGGAGGAGGAGCAGAAAGCAATGAGAGAAA	300		
QY	493	TCCTTGGGAAAAACACAAACCGAACTTACTAAGAGAGTGAAGACAGTGTAGAGTTTGC	552		
Db	301	TCCTTGGGAAAAACACAAACCGAACTTACTAAGAGAGTGAAGACAGTGTAGAGTTTGC	360		



```

QY 553 TAAACACATGATTAAGTGGCCATGCTGTGTACTGGACCCAGCCCTGCCAGCCCAATCA 612
|
|
|
Db 361 TAAACACATGATTAAGTGGCCATGCTGTGTACTGGACCCAGCCCTGCCAGCCCAATCA 420
|
|
|
QY 613 CCCTGAAGGAGTGGCCATGCTGTGTACTGGACCCAGCCCTGCCAGCCCAATCA 672
|
|
|
Db 421 CCCTGAAGGAGTGGCCATGCTGTGTACTGGACCCAGCCCTGCCAGCCCAATCA 480
|
|
|
QY 673 TTGCAAAACCAAAAACCCCTCTACCAATGCTCAGTGTGTGTGTGTGTGTGTGTGTGT 732
|
|
|
Db 481 TTGCAAAACCAAAAACCCCTCTACCAATGCTCAGTGTGTGTGTGTGTGTGTGTGTGT 540
|
|
|
QY 733 ATGTTGAAGTGGCCAGTCTGCTGCTTAAAGGAATAGCAGACACAGTATGGAGGG 792
|
|
|
Db 541 ATGTTGAAGTGGCCAGTCTGCTGCTTAAAGGAATAGCAGACACAGTATGGAGGG 600
|
|
|
QY 793 GGTACTTTGACAAACAGATAAAGAGTCCCTGGATTAAGTGTGCTAATCAAGATTTGGCTTTC 852
|
|
|
Db 601 GCTACTTTGACAAACAGATAAAGAGTCCCTGGATTAAGTGTGCTAATCAAGATTTGGCTTTC 660
|
|
|
QY 853 TGAGTGGATTTCCACCAACATTCAGGAACAAATCACTTAAACAAATTCGCAGAAA 912
|
|
|
Db 661 TGAGTGGATTTCCACCAACATTCAGGAACAAATCACTTAAACAAATTCGCAGAAA 720
|
|
|
QY 913 ATCTTTCTCTAAAGTCAAACTGTTGAGTGTGTATTCAGAGCCAGTGTAGCGCTTCTAA 972
|
|
|
Db 721 ATCTTTCTCTAAAGTCAAACTGTTGAGTGTGTATTCAGAGCCAGTGTAGCGCTTCTAA 780
|
|
|
QY 973 AACTGAATAACCCCAAGATTTCAAGATTTGAATAG-CAAACTAAGAAGACATGACC 1031
|
|
|
Db 781 AACTGAATAACCCCAAGATTTCAAGATTTGAATAGCCAACTAAGAAGACATGACC 840
|
|
|
QY 1032 ATTG-ATGGAAGAAGTCAAACTGACCAATAGTCCCTGCATATTTAT 1074
|
|
|
Db 841 ATTGNATGGGAAGACATGACCCATAGTCCCTGCATATTTAT 884
|
|
|

RESULT 3
BM476404
LOCUS
DEFINITION
AGENCOURT 6476164 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5553363
5', mRNA sequence.
BM476404
ACCESSION
VERSION
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 994)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-research.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LIAW12270 row: d column: 04
High quality sequence stop: 676.
Location/Qualifiers
1. .994
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5553363"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_85"

```

## FEATURES

## source

## ORIGIN

```

Query Match      14.5%; Score 813; DB 4; Length 994;
Best Local Similarity 95.9%; Pred. No. 2.8e-120;
Matches 910; Conservative 0; Mismatches 30; Indels 9; Gaps 7;

QY 4566 CATCTGAATGCTTTGTAATTTGACCTTCAGCCCTTAAGATGCTATGATTTTCAGTGC 4625
|
|
|
Db 1 CATCTGAATGCTTTGTAATTTGACCTTCAGCCCTTAAGATGCTATGATTTTCAGTGC 60
|
|
|
QY 4626 AGGTCTAATTTCAACAGGCTAGAGTACTTACCTACAGATGA-ATTATGTTTGA 4684
|
|
|
Db 61 AGGTCTAATTTCAACAGGCTAGAGTACTTACCTACAGATGANATTATGTTTGA 120
|
|
|
QY 4685 AATGTACATATTCAACAGAGTGGCTCATTTTAGAATGAGTACTGCTGATGGCACTGG 4744
|
|
|
Db 121 AATGTACATATTCAACAGAGTGGCTCATTTTAGAATGAGTACTGCTGATGGCACTGG 180
|
|
|
QY 4745 CACATTACAGTGGTGTCTTTGTTAATTAATCTCAATTTGTTTCCAGTAGCTATCTCTCA 4804
|
|
|
Db 181 CACATTACAGTGGTGTCTTTGTTAATTAATCTCAATTTGTTTCCAGTAGCTATCTCTCA 240
|
|
|
QY 4805 GTTGGTTTTGATAGACAGAGCCAGCAAACTTTCTTTGTTAAAGGCTGTAGTAAT 4864
|
|
|
Db 241 GTTGGTTTTGATAGACAGAGCCAGCAAACTTTCTTTGTTAAAGGCTGTAGTAAT 300
|
|
|
QY 4865 TATTGAGGCCACCTGTGTCTTTGTTGATACATTTCTTCTGTTGTTTGTAGTTGTTT 4924
|
|
|
Db 301 TATTGAGGCCACCTGTGTCTTTGTTGATACATTTCTTCTGTTGTTTGTAGTTGTTT 360
|
|
|
QY 4925 TTTCAAAACAAACCTCTAAAAATGTAAACCAATGTTTGTAGCTGTGTACAAAACCTG 4984
|
|
|
Db 361 TTTCAAAACAAACCTCTAAAAATGTAAACCAATGTTTGTAGCTGTGTACAAAACCTG 420
|
|
|
QY 4985 CCCACAGCCAGATGTGACCTTCAGCCCATATTTGCCAATCACTGAGATTA-TTTTTG 5043
|
|
|
Db 421 CCCACAGCCAGATGTGACCTTCAGCCCATATTTGCCAATCACTGAGATTAAGTTT 480
|
|
|
QY 5044 TTGTTGTTGTTGTTGTTGTTGTTGATAGACAGAGTCTCTCTGTTGCCAGGCTGAGTGC 5103
|
|
|
Db 481 TTGTTGTTGTTGTTGTTGTTGTTGATAGACAGAGTCTCTCTGTTGCCAGGCTGAGTGC 540
|
|
|
QY 5104 AGTGGCGCAATCTCAGCTCACTGCAACCTCCGCTCCGCGGTTCAGAGCACTCTCTCA 5163
|
|
|
Db 541 AGTGGCGCAATCTCAGCTCACTGCAACCTCCGCTCCGCGGTTCAGAGCACTCTCTCA 600
|
|
|
QY 5164 GCCTTCTGAGTAGCTGGGACTACAGGTGCATGCCACCAACCCCTGCTAATTTTGTATTT 5223
|
|
|
Db 601 GCCTTCTGAGTAGCTGGGACTACAGGTGCATGCCACCAACCCCTGCTAATTTTGTATTT 660
|
|
|
QY 5224 TTATAGACAGCGGGGTTCCACATATTTGGTCAAGCTTATCTTGAATCTCAGCTCAGG 5283
|
|
|
Db 661 TTATAGACAGCGGGGTTCCACATATTTGGTCAAGCTTATCTTGAATCTCAGCTCAGG 720
|
|
|
QY 5284 TGATCCACTGCTCTGCTCCCAAGTGTGAGATTAAGGCAATAGCCAGTGCACCCA 5343
|
|
|
Db 721 TGATCCACTGCTCTGCTCCCAAGTGTGAGATTAAGGCAATAGCCAGTGCACCCA 780
|
|
|
QY 5344 GCCGAGAATTAGTATTTTATGATGTTTAA-CTTGGCGTCTAGCCATATTTATGTC 5402
|
|
|
Db 781 GCCGAAAATAAAGATTTTATGATGTTTAA-CTTGGCGTCTAGCCATATTTATGTC 840
|
|
|
QY 5403 ATAATACATGGA-TTTGTGAGAGCAGATTCATGAGTAA-CTCTGACAGTA--TTTT 5458
|
|
|
Db 841 ATAATACATGGA-TTTGTGAGAGCAGATTCATGAGTAA-CTCTGACAGTA--TTTT 900
|
|
|
QY 5459 AGATCATGATCTCAACAATATTCCT--CCCAATGGCATACATCTTTT 5505
|
|
|

```

```

/Note="Organ: lymph; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.867 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

```

Db 901 AACCTGGATCTCCACATATTTCTTCCCAAAAGGGGATACATCTTTTG 949

RESULT 4  
BQ053143  
LOCUS  
DEFINITION BQ053143 981 bp mRNA linear EST 29-MAR-2002  
AGENCOURT 6821868 NIH\_MGC\_106 Homo sapiens cDNA clone IMAGE:5934740  
5' mRNA sequence.  
ACCESSION BQ053143  
VERSION BQ053143.1 GI:19812483  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 981)  
NIH-MGC http://mhc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-r@mail.nih.gov  
Tissue Procurement: Dr. Daniel McVicar, DBS/NCI  
cDNA Library Preparation: Rubin Laboratory  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: Agencourt Bioscience Corporation  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L10M1211 row: f column: 21  
High quality sequence stop: 575.  
Location/Qualifiers  
1. 981  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5934740"  
/tissue\_type="natural killer cells, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 106"  
/note="Organ: blood; Vector: pOTB7; Site: 1: XhoI; Site 2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGG(G). Library constructed by Ling Hong in the  
Laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH\_MGC Library."

ORIGIN  
Query Match 14.58; Score 812; DB 5; Length 981;  
Best Local Similarity 95.48; Pred. No. 4e-120;  
Matches 867; Conservative 0; Mismatches 38; Indels 4; Gaps 3;

Qy 3315 GAACACATGCTAATAAATGCTCAAAAATGCTTCCTGCTGATATACGAGCTAAATAAT 3374  
Db 1 GAACACATGCTAATAAATGCTCAAAAATGCTTCCTGCTGATATACGAGCTAAATAAT 60

Qy 3375 ATTCCACCACTAGGAATCCTACTATGATCCCACTGCGCCGCTCACTAAAGTCTA 3434  
Db 61 ATTCCACCACTAGGAATCCTACTATGATCCCACTGCGCCGCTCACTAAAGTCTA 120

Qy 3435 GTACCAAACTGTAACACAGTAACCTGACAAAATCCAAAGCATATAGGACAAACAAA 3494  
Db 121 GTACCAAACTGTAACACAGTAACCTGACAAAATCCAAAGCATATAGGACAAACAAA 180

Qy 3495 TATAGTTTGAATCATGAGGAGAGAAATCGTTTTAAATGATGTTACCAACGTT 3554  
Db 181 TATAGTTTGAATCATGAGGAGAGAAATCGTTTTAAATGATGTTACCAACGTT 240

Qy 3555 TCTCATGTGTTGCCAGTTGGATGACATAGAAAACCTAGGAAGTTTGTTCCTG 3614  
Db 241 TCTCATGTGTTGCCAGTTGGATGACATAGAAAACCTAGGAAGTTTGTTCCTG 300

Qy 3615 AATGACAACATTGACCACATATATAAGATGCTCAGACAGTGAAGGCTGTTCTCAGGGAC 3674  
Db 301 AATGACAACATTGACCACATATATAAGATGCTCAGACAGTGAAGGCTGTTCTCAGGGAC 360

Qy 3675 TTCATGAATCCATGTTCCCATACCTTCCCAATTTGAATTCGCAAGAGAGTATCGAAAC 3734  
Db 361 TTCATGAATCCATGTTCCCATACCTTCCCAATTTGAATTCGCAAGAGAGTATCGAAAC 420

Qy 3735 CGTTTCCTTCATATGATGAGCTGCAGCAATGGAGGCTTATCGAGACAAATTTGAAGTTT 3794  
Db 421 CGTTTCCTTCATATGATGAGCTGCAGCAATGGAGGCTTATCGAGACAAATTTGAAGTTT 480

Qy 3795 TGGACCCATTTGTACTAGCAACATTCATTATGTTTACTATATTTCTCATTTTGTCTGAG 3854  
Db 481 TGGACCCATTTGTACTAGCAACATTCATTATGTTTACTATATTTCTCATTTTGTCTGAG 540

Qy 3855 CAGTTAATTCACCTTAAGCGGAAGATATTTCCAGAGAGGAGTACACAAAGAGCTAGT 3914  
Db 541 CAGTTAATTCACCTTAAGCGGAAGATATTTCCAGAGAGGAGTACACAAAGAGCTAGT 600

Qy 3915 CCCAATCGAATCAGAGTATAGAGATCTTCATTTTGAACCATCTACCTCAGCATTTACT 3974  
Db 601 CCCAATCGAATCAGAGTATAGAGATCTTCATTTTGAACCATCTACCTCAGCATTTACT 660

Qy 3975 GAGCATTTTAAACTCAGTTTCAGAGATGCTTTTGTGATGATGCTTAGCAG-TTTG 4033  
Db 661 GAGCATTTTAAACTCAGTTTCAGAGATGCTTTTGTGATGATGCTTAGCAG-TTTG 720

Qy 4034 GCCCGAGAAGGAAATATCCAGTACCATGCTGTTTTGTGCGATGAATAGCCCACTGA 4093  
Db 721 CCCCGAGAAGGAAATATCCAGTACCATGCTGTTTTGTGCGATGAATAGCCCACTGA 780

Qy 4094 CTAGGAATTTTAAACCAACCCACTG-AAACTTGTGTGCGAGCACTCTGAACCTGATT 4152  
Db 781 CCAGGAATTTTAAACCAACCCACTGAAACCTGTTGTTTTCAGAGCTCTGAACCTGATT 840

Qy 4153 TTACTTTTAAAGATTTT-GCTCATGACCTGTCATCTTTTATAAAAGGCTCACTGA 4210  
Db 841 TTACTTTTAAAGATTTTGTCTTCAAGCCGTCATCTTTTATAAAAGGCTCACTGA 900

Qy 4211 CAAGAGACA 4219  
Db 901 CCTGACAAA 909

RESULT 5  
BQ452210  
LOCUS  
DEFINITION BQ452210 1118 bp mRNA linear EST 05-FEB-2002  
AGENCOURT 6386314 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:5526571  
5' mRNA sequence.  
ACCESSION BQ452210  
VERSION BQ452210.1 GI:18501250  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1118)  
NIH-MGC http://mhc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-r@mail.nih.gov  
Tissue Procurement: ATCC/BCTD/DTF  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: Agencourt Bioscience Corporation  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L10M12200 row: g column: 20  
High quality sequence stop: 717.  
Location/Qualifiers

FEATURES



extraction, and separated from free linkers by Centricon-100 column. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer SalI-S for 25 cycles. The products were purified by phenol/chloroform extraction and Centricon-100 column. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of PCMV-SPORT6 plasmid vector. The average insert size is about 3.6kb."

ORIGIN		Query Match		Best Local Similarity		Matches		Conservative		0; Mismatches		15; Indels		1; Gaps	
2616		GTGAAACTCACCACCCCAAAACCATAGCGGGAATGTGACAAAAGAAAGCCCATCT		2675		12		GTGAAACTCACCACCCCAAAACCATAGCGGGAATGTGACAAAAGAAAGCCCATCT		71					
2676		CTGATTCTCCACTGGAAACCCAGATGACAAAAGAAAGAAATCACAGGAAAGAAAA		2735		72		CTGATTCTCCACTGGAAACCCAGATGACAAAAGAAAGAAATCACAGGAAAGAAAA		131					
2736		GAGAACAGTAGAATGGAGGAAATGCTGAAATACATAGCGGCTTACTGAAAGTTACTT		2795		132		GAGAACAGTAGAATGGAGGAAATGCTGAAATACATAGCGGCTTACTGAAAGTTACTT		191					
2796		GGAAGAACTGCAGCATACAGATAGTACTTGGCTTTTGGCATGGGAGAAAAA		2855		192		GGAAGAACTGCAGCATACAGATAGTACTTGGCTTTTGGCATGGGAGAAAAA		251					
2856		AAGTATTTCACAGATCTTCTCGACGAGAAAGAGTCATTGGAACACAAATGGCATCTTC		2915		252		AAGTATTTCACAGATCTTCTCGACGAGAAAGAGTCATTGGAACACAAATGGCATCTTC		311					
2916		ACTGATGACAAAATATCTGGGACCACTAAAGATACATTTCCAGATCCCTCAGATAT		2975		312		ACTGATGACAAAATATCTGGGACCACTAAAGATACATTTCCAGATCCCTCAGATAT		371					
2976		GTAATAAAATTTCAATAGCAAGTTTGGATTACATCGCGGAAAGTCCCTGCTCATG		3035		372		GTAATAAAATTTCAATAGCAAGTTTGGATTACATCGCGGAAAGTCCCTGCTCATG		431					
3036		CCTCAGATGATGACCGGATGTTATGCAAGAACTGCAAGATATGTTCCCTGAGAAATT		3095		432		CCTCAGATGATGACCGGATGTTATGCAAGAACTGCAAGATATGTTCCCTGAGAAATT		491					
3096		GACAGAGCTCATTTCAAAAGTCGCGCATTTGAGGATATGAGATTTGCTTTCTCTAT		3155		492		GACAGAGCTCATTTCAAAAGTCGCGCATTTGAGGATATGAGATTTGCTTTCTCTAT		551					
3156		TTTTATTATCTCATGAGTGCAGCCACTGAATATATCTCAAGTCTTTGATGAGTT		3215		552		TTTTATTATCTCATGAGTGCAGCCACTGAATATATCTCAAGTCTTTGATGAGTT		611					
3216		GATACAGATCAATCTGCTGTTCTGTCGACAGAGAAATCCGACACTGGCTACCAAGTT		3275		612		GATACAGATCAATCTGCTGTTCTGTCGACAGAGAAATCCGACACTGGCTACCAAGTT		671					
3276		CAGCAATCTGCGTTAAGTTTTCAGGATTTGACAGGCTGGAACACATGCTTAATAATTC		3335		672		CAGCAATCTGCGTTAAGTTTTCAGGATTTGACAGGCTGGAACACATGCTTAATAATTC		731					
3336		TCAAAAATGCTTCTGCTGATATCAGCAGCTAAATTAATTTCCACCACTCAGGATCC		3395		732		TCAAAAATGCTTCTGCTGATATCAGCAGCTAAATTAATTTCCACCACTCAGGATCC		791					
3396		TACTATGATCCCAACCTGCCACCGGTCACTAAAAGTCTAGTAACAA		3441		792		TACTATGATCCCAACCTGCCACCGGTCACTAAAAGTCTAGTAACAA		836					

BM450679  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source

BM450679  
AGENCOURT\_6394543 NIH\_MGC\_67 Homo sapiens cDNA clone IMAGE:5494379  
5', mRNA sequence.  
BM450679  
BM450679.1 GI:18499719  
EST.  
Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1061)  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: csapbs@mail.nih.gov  
Tissue Procurement: ATCC  
DNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1M12119 row: j column: 12  
High quality sequence stop: 637.  
Location/Qualifiers  
1..1061  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5494379"  
/tissue\_type="retinoblastoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_67"  
/note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 1.75 kb. Library constructed by Life  
Technologies."

ORIGIN		Query Match		Best Local Similarity		Matches		Conservative		0; Mismatches		27; Indels		8; Gaps	
640		CTTCTTTTCATCTGCCAGTGCACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCA		699		25		CTTCTTTTCATCTGCCAGTGCACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCA		84					
700		ATGCTCTCAGTTGTTGTTTTCACAGTACTAAGGATGTTGAAGATGCCCACTCTGCACTGC		759		85		ATGCTCTCAGTTGTTGTTTTCACAGTACTAAGGATGTTGAAGATGCCCACTCTGCACTGC		144					
760		TTAAAGGAAATAGCAGACAGACAGATGAGAGGGGTACTTCACACAGATTAAGAGTCC		819		145		TTAAAGGAAATAGCAGACAGACAGATGAGAGGGGTACTTCACACAGATTAAGAGTCC		204					
820		CTGGATTAGTCTAATCAAGATTTGGCTTTCTCCTGAGTGGATTTCCCAACATTTCAAGG		879		205		CTGGATTAGTCTAATCAAGATTTGGCTTTCTCCTGAGTGGATTTCCCAACATTTCAAGG		264					
880		AAACAAATCAACTAAAAACAAAATTCAGAAAATCTTTTCTCTAAAGTCAAACTGTTC		939		265		AAACAAATCAACTAAAAACAAAATTCAGAAAATCTTTTCTCTAAAGTCAAACTGTTC		324					
940		AGTTGATTTACAGGCGCTGTAGCGCTTCTTAAACTGAATAACCCCAAGGATTTTCAAG		999		325		AGTTGATTTACAGGCGCTGTAGCGCTTCTTAAACTGAATAACCCCAAGGATTTTCAAG		384					
1000		AATTGAATAAGCAAACTAAGAGAACATGACCATTTGATGGAAAAAGAACTGACCAATGTC		1059		385		AATTGAATAAGCAAACTAAGAGAACATGACCATTTGATGGAAAAAGAACTGACCAATGTC		444					



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 867)  
 NIH-MGC <http://mgc.ncl.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Daniela S. Gerhard, Ph.D.  
 Office of Cancer Genomics / NIH  
 National Cancer Institute / NIH  
 Bldg. 31 Rm10A07 Bethesda, MD 20892  
 Email: cgabs-remail.nih.gov  
 Tissue Procurement: Irene Ginis and Mahendra Rao, NIA  
 cDNA Library Preparation: Yulan Piao and Minoru Ko  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC C lone distribution information  
 can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: NDAM495 row: h column: 06  
 High quality sequence stop: 657.  
 Location/Qualifiers  
 1..867  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:30419981"  
 /tissue\_type="Embryonic Stem cells"  
 /cell\_line="WA01"  
 /lab\_hosts="DH10B (T1 phage-resistant)"  
 /clone\_lib="NIA Human H1 Embryonic Stem Cell cDNA Library (Long)"  
 /notes="vector: pCMV-Sport6; Site 1: NotI; Site 2: SalI; This is a long-transcript enriched cDNA library (Genome Res. 11: 1553-1558 (2001)). [PMID: 11544199] from WA01 cell line. Undifferentiated human ES cell line WA01/H1 was obtained from WiCell Research Institute, Inc., Madison, WI, cultured according to their instructions, on MEF feeders. They formed round colonies with defined edges and were positive for alkaline phosphatase, SSEA-4, OCT3, OCT4, REX1, UTF, TERT, SOX2, CX43 and CX45. They are negative for GATA2, GATA4, PDX1, NCAM, MSX1, FLT3, SSEA-1, TUBB3, NES, GFAP, and BOMES. When confluent (18-10 days after plating), the ES cells from 4 x 6 cm dishes were treated with 1 mg/ml collagenase, type IV (Invitrogen/GIBCO) for 5-10 min and gently scraped off with 5 ml pipette. RNA was purified with TRIzol Reagent from Invitrogen. Protocol ref: Genome Res. 11: 1553-1558 (2001). [PMID:11544199] Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen: 5'-pGATCTTCTAGATCCGAGCGCGCCCTTTTCTTTT-3'] from 3.4g of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker Ls-Sal4, purified by phenol/chloroform extraction, and separated from free linkers by Centricon-100 column. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S for 25 cycles. The products were purified by phenol/chloroform extraction and Centricon-100 column. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The average insert size is about 3.6kb."

#### FEATURES

source

Db 71 CTGATTCTTCCACTGGAAAGCCAGATGACAAAAAGAAAAATCACAGGGAAGAAAA 130  
 QY 2736 GAGAACAGTAGAATGGAGGAAATGCTGAAATCAACATAGGCGCTTACTGAAGTGTACTT 2795  
 Db 131 GAGAACAGTAGAATGGAGGAAATGCTGAAATCAACATAGGCGCTTACTGAAGTGTACTT 190  
 QY 2796 GGAAGAAAGCTGCAGCATTACACAGATAGTACTTGGGCTTTTGGCATGGGAGAAAAA 2855  
 Db 191 GGAAGAAAGCTGCAGCATTACACAGATAGTACTTGGGCTTTTGGCATGGGAGAAAAA 250  
 QY 2856 AAGTATTTTCAAGATCTTCTCGAGGAAGAGAGTCAATGGAAGACACAATTTGGCATACTTC 2915  
 Db 251 AAGTATTTTCAAGATCTTCTCGAGGAAGAGAGTCAATGGAAGACACAATTTGGCATACTTC 310  
 QY 2916 ACTGATAGCAAAATATCTGGAGGCAACTAAAGATACATTTGCAGATTCCTCAGATAT 2975  
 Db 311 ACTGATAGCAAAATATCTGGAGGCAACTAAAGATACATTTGCAGATTCCTCAGATAT 370  
 QY 2976 GTAATATAATTTCTAAATAGCAAGTTTGGATTCACATCGCGGAAGTCCCTGCTCACATG 3035  
 Db 371 GTAATATAATTTCTAAATAGCAAGTTTGGATTCACATCGCGGAAGTCCCTGCTCACATG 430  
 QY 3036 CCTCACATGATTGACCGGATTGTTATGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTT 3095  
 Db 431 CCTCACATGATTGACCGGATTGTTATGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTT 490  
 QY 3096 GACAAGACGTCATTTCAAAAGTGGCCCATCTGAGGATATGAGTATGAGTATGAGTATGAGTAT 3155  
 Db 491 GACAAGACGTCATTTCAAAAGTGGCCCATCTCTGAGGATATGAGTATGAGTATGAGTAT 550  
 QY 3156 TTTTATTATCTCATGAGTGGAGTGCAGCCACTGAAATATATCTCAAGTCTTTGATGAAGTT 3215  
 Db 551 TTTTATTATCTCATGAGTGGAGTGCAGCCACTGAAATATATCTCAAGTCTTTGATGAAGTT 610  
 QY 3216 GATACAGATCAATCTGCTGCTTCTGTCAGAGAGAAATCCGAACACCTGGCTACAGAAAT 3275  
 Db 611 GATACAGATCAATCTGCTGCTTCTGTCAGAGAGAAATCCGAACACCTGGCTACAGAAAT 669  
 QY 3276 CACGAACCTGCGTTAACTTTGCGAGATTTGCGAGTCTGGAACACATGCTTAATAAATGTC 3335  
 Db 670 CACGAACCTGCGGCTTAGTTTTCAGATGTCAGATGTCAGATGTCAGATGTCAGATGTCAGAT 729  
 QY 3336 TCAAAAATGCTTCTGCTGATATCACGAGCTAAATAATATTTCCCAACTCAGGAATCC 3395  
 Db 730 CTCAAANNCTCTGCTGATATCACGAGCTAAATAATATTTCCCAACTCAGGAATCC 789  
 QY 3396 TACTATGATCCCAACTGCT 3414  
 Db 790 TACTATGATCCNNACTGC 808  
 RESULT 10  
 BQ720582  
 LOCUS  
 DEFINITION  
 IMAGE:6193590 5', mRNA sequence.  
 ACCESSION  
 BQ720582  
 VERSION  
 EST.  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 929)  
 NIH-MGC <http://mgc.ncl.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: Dr. James R. Lupski  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

#### ORIGIN

Query Match 13.4%; Score 751.8; DB 6; Length 867;  
 Best Local Similarity 96.7%; Pred. No. 1.8e-110;  
 Matches 773; Conservative 0; Mismatches 25; Indels 1; Gaps 1;  
 QY 2616 GTGGAACTCACCCAAAAACCATAGCGGAATGTGACAAAAGAAAAAGCCCATCT 2675  
 Db 11 GTGGAACTCACCCAAAAACCATAGCGGAATGTGACAAAAGAAAAAGCCCATCT 70  
 QY 2676 CTGATTGTTCCACTGGAAAGCCAGATGACAAAAGAAAAATCACAGGGAAGAAAA 2735



DNA sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LRAM13597 row: h column: 07  
High quality sequence stop: 693.

## FEATURES

Location/Qualifiers  
1..929  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6193590"  
/sex="male"  
/tissue\_type="sympathetic trunk"  
/dev\_stage="adult, 16 yr"  
/lab\_host="DH10B"  
/clone\_lib="Lupski sympathetic trunk"  
/note="Vector: pCMV-SPORT6 (Life Technologies); Site 1:  
NotI; Site 2: SalI; cDNA made by oligo-dT priming.  
Directionally cloned using the following adaptors:  
5'-TCGACCCACGCGTCCG-3' and  
5'-CACTAGTCTAGATCGGAGCGCGCCCT(15)-3'. Size selected >  
1 kb for average insert length 1.9 kb. This is a primary  
library, non-amplified. Library constructed by Life  
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor  
College of Medicine); available through Life  
Technologies."

## ORIGIN

Query Match 13.3%; Score 746.8; DB 5; Length 929;  
Best Local Similarity 93.8%; Pred. No. 1.1e-109;  
Matches 845; Conservative 0; Mismatches 43; Indels 13; Gaps 6;  
Db 1 TGGACACATGCTAATAATGCTCAAAATGCTTCTGCTGATATCAGCAGCTAATA 3372  
Qy 3313 TGGACACATGCTAATAATGCTCAAAATGCTTCTGCTGATATCAGCAGCTAATA 3372  
Db 1 TGGACACATGCTAATAATGCTCAAAATGCTTCTGCTGATATCAGCAGCTAATA 60  
Qy 3373 ATATCCACCAACTCAGGAATCTTACTATGATCCCAACTGCCACCGGTCACTAAAGTC 3432  
Db 61 ATATCCACCAACTCAGGAATCTTACTATGATCCCAACTGCCACCGGTCACTAAAGTC 120  
Qy 3433 TAGTACAAACTGTAAACCACTACTGACAAATCCCAAGCATATAGGACAAACA 3492  
Db 121 TAGTACAAACTGTAAACCACTACTGACAAATCCCAAGCATATAGGACAAACA 180  
Qy 3493 AATATAGTGTGAATCATGGAGAGAGAAATCGCTTTTAAATGATTCGTACCAACG 3552  
Db 181 AATATAGTGTGAATCATGGAGAGAGAAATCGCTTTTAAATGATTCGTACCAACG 240  
Qy 3553 TTTCTCATGTGGTGGCCAGTGGATGACATAGAAAAACCCCTAGGAAGTTTGTGGCC 3612  
Db 241 TTTCTCATGTGGTGGCCAGTGGATGACATAGAAAAACCCCTAGGAAGTTTGTGGCC 300  
Qy 3613 TGAATGACAACTTGACCAATCATAAAGATGCTCAGACAGTGAAGGCTGTCTCAGGG 3672  
Db 301 TGAATGACAACTTGACCAATCATAAAGATGCTCAGACAGTGAAGGCTGTCTCAGGG 360  
Qy 3673 ACTTCTATGAATCATGTGCCATACCTTCCCAATTTGAACCTGCCAAGAGATGCGAA 3732  
Db 361 ACTTCTATGAATCATGTGCCATACCTTCCCAATTTGAACCTGCCAAGAGATGCGAA 420  
Qy 3733 ACCGTTTCCCTCATATGATGAGCTGAGGAGTGGAGGCTTATCCGACAAATTTGAAT 3792  
Db 421 ACCGTTTCCCTCATATGATGAGCTGAGGAGTGGAGGCTTATCCGACAAATTTGAAT 480  
Qy 3793 TTTGGACCAATTTGTACTAGCAACATTTATGTTTACTATATTTCTCAATTTTGTGCTG 3852  
Db 481 TTTGGACCAATTTGTACTAGCAACATTTATGTTTACTATATTTCTCAATTTTGTGCTG 540  
Qy 3853 AGCAGTTAATGCACTTAAGCGGAGATATTTCCCAAGGAGATACACAAAGAGCTA 3912  
Db 541 AGCAGTTAATGCACTTAAGCGGAGATATTTCCCAAGGAGATACACAAAGAGCTA 600

Qy 3913 GTCCCAATCGAATCAGATATAGAGATCTTCAATTTGAAAAACCATCTACTCAGCATTTA 3972  
Db 601 GTCCCAATCGAATCAGATATAGAGATCTTCAATTTGAAAAACCATCTACTCAGCATTTA 660  
Qy 3973 CTGAGCATTTTAAACTCAGCTTTCACAGAGATGCTTTCTGTGATGCA---TGCTTAGCAG 4029  
Db 661 CTGAGTATTTTAAACTCAGCTTTCACAGAGATGCTTTCTGTGATGCA---TGCTTAGCAG 720  
Qy 4030 TTGGCCCGAAGAGAAATATCCAGTACC---ATGCTGTTTGTGGCATGAATATAGC 4086  
Db 721 TTGGCCCGAAGAGAAATATCCAGTACCAGTACCCATGCTTTTGTGGCATGAATATAGC 780  
Qy 4087 CCACCTGACTA-GGATTTATTAACCAACCACTGAAAACCTCTGT---GTCGAGCAGCTC 4142  
Db 781 CCACTGACCGAGGATTTATTAACCAACCAACCGGAAAACCTTTGTGTTGGTTGAGCAGCTC 840  
Qy 4143 TG-AACTGATTTTACTTTTAAAGAAATTTG--CTCATGGACCTGTCTCATCTTTTATAAA 4199  
Db 841 TGAACATGATTTTACTTTTAAAGAAATTTGCTTGGCTCTGGGACTTGTCTATCCCTTTTATAA 900  
Qy 4200 A 4200  
Db 901 A 901  
RESULT 11  
BG214021  
LOCUS BG214021 783 bp mRNA linear EST 21-APR-2001  
DEFINITION RST33648 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.  
ACCESSION BG214021  
VERSION BG214021.1 GI:13735708  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 783)  
AUTHORS Harrington,J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,  
Cain,S., Leventhal,C., Thornton,M., Ramachandran,R.,  
Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Booser,S.,  
Mays,R., Smith,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K.,  
Offenbacher,J., Danig,J. and Ducar,W.  
TITLE Creation of genome-wide protein expression libraries using random  
activation of genome-wide protein expression libraries using random  
Nat. Biotechnol. 19 (5), 440-445 (2001)  
JOURNAL MEDLINE 21227151  
PURVED 11329013  
COMMENT Contact: Scott J. Cain  
Athersys, Inc.  
3201 Carnegie Ave, Cleveland, OH 44115, USA  
Tel: 216 431 9900  
Fax: 216 361 9596  
Email: [scain@athersys.com](mailto:scain@athersys.com)  
High quality sequence stop: 522.  
Location/Qualifiers  
1..783  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/cell\_line="HT1080"  
/clone\_lib="Athersys RAGE Library"  
/note="See 'Creation of Genome-wide Protein Expression  
Libraries using Random Activation of Gene Expression',  
Nature Biotechnology, in press. Note that even though the  
cell type indicated is HT1080, since a random activation  
method was used, these sequence tags are not necessarily  
expressed in HT1080 under normal circumstances."

## ORIGIN

Query Match 13.3%; Score 743.8; DB 4; Length 783;  
Best Local Similarity 98.7%; Pred. No. 3.5e-109;  
Matches 771; Conservative 0; Mismatches 7; Indels 3; Gaps 2;

3765 TGGGGGCTTATCGAGCAAAATTGAAGTTTGGACCCATTGTGTACTAGCAATGATT 3824  
 Db 3 TGGGGGCTTATCGAGCAAAATTGAAGTTTGGACCCATTGTGTACTAGCAATGATT 61  
 3825 ATGTTTACTATTTCTCAATTTTGTGTCAGCAGTAAATGCACTTAAGCGGAAGATATTT 3884  
 Db 62 ATGTTTACTATTTCTCAATTTTGTGTCAGCAGTAAATGCACTTAAGCGGAAGATATTT 121  
 3885 CCAGAGAGGAGTACACAAAGAGCTAGTCCCAATCGAATCAGATATAGAGATCTTC 3944  
 Db 122 CCAGAGAGGAGTACACAAAGAGCTAGTCCCAATCGAATCAGATATAGAGATCTTC 181  
 3945 ATTTGAAACCATCTACTCAGCATTTTCTGAGCATTTTAAACTCAGCTTCACAGAT 4004  
 Db 182 ATTTGAAACCGTCTACTCAGCATTTTCTGAGCATTTTAAACTCAGCTTCACAGAT 241  
 4005 GTCTTTGTGATGTAGCTTTAGCAGTTTGGCCGAGAGAGAAATATCCAGTACCATGC 4064  
 Db 242 GTCTTTGTGATGTAGCTTTAGCAGTTTGGCCGAGAGAGAAATATCCAGTACCATGC 301  
 4065 TGTCTTGTGATGTAGCTTTAGCAGTTTGGCCGAGAGAGAAATATCCAGTACCATGC 4124  
 Db 302 TGTCTTGTGATGTAGCTTTAGCAGTTTGGCCGAGAGAGAAATATCCAGTACCATGC 361  
 4125 TTGTGTCTGAGCAGCTCTGAACTGATTTTCTTAAAGAAATTTGCTCATGACCTGTC 4184  
 Db 362 TTGTGTCTGAGCAGCTCTGAACTGATTTTCTTAAAGAAATTTGCTCATGACCTGTC 421  
 4185 ATCTTTTAAAGAGCTCTGAACTGATTTTCTTAAAGAAATTTGCTCATGACCTGTC 4244  
 Db 422 ATCTTTTAAAGAGCTCTGAACTGATTTTCTTAAAGAAATTTGCTCATGACCTGTC 481  
 4245 TGCAGACTAATCTTATAGAGAGCTCTGAACTGATTTTCTTAAAGAAATTTGCTCATGACCTGTC 4304  
 Db 482 TGCAGACTAATCTTATAGAGAGCTCTGAACTGATTTTCTTAAAGAAATTTGCTCATGACCTGTC 541  
 4305 AGTCTTTGATCTCAAGAGCTTTGCTTAAAGAAATTTGCTCATGACCTGTC 4362  
 Db 542 AGTCTTTGATCTCAAGAGCTTTGCTTAAAGAAATTTGCTCATGACCTGTC 601  
 4363 ATTTTAAAGAGCTTTGCTTAAAGAAATTTGCTCATGACCTGTC 4422  
 Db 602 ATTTTAAAGAGCTTTGCTTAAAGAAATTTGCTCATGACCTGTC 661  
 4423 TGTCTAAACCTATTTTATAGATGTTTATTAATATGAGCTGTAATGCTTAAACCTAA 4482  
 Db 662 TGTCTAAACCTATTTTATAGATGTTTATTAATATGAGCTGTAATGCTTAAACCTAA 721  
 4483 CAATACCTAATATGAGTATTTATTAATATGAGCTGTAATGCTTAAACCTAA 4542  
 Db 722 CAATACCTAATATGAGTATTTATTAATATGAGCTGTAATGCTTAAACCTAA 781  
 4543 C 4543  
 Db 782 C 782

RESULT 12  
 CD652380 819 bp mRNA linear EST 18-JUN-2003  
 LOCUS AGNCCOURT 14537840 NIA Human H1 Embryonic Stem Cell cDNA Library  
 DEFINITION (Long) Homo sapiens cDNA clone IMAGE:30421320 5', mRNA sequence.  
 ACCESSION CD652380  
 VERSION CD652380.1 GI:31889706  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 819)  
 NH-MGC http://mgi.nci.nih.gov/  
 TITILE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)

COMMENT

Contact: Daniela S. Gerhard, Ph.D.  
 Office of Cancer Genomics / NIH  
 Bldg. 31 Rm10A07 Bethesda, MD 20892  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Irene Ginis and Mahendra Rao, NIA  
 cDNA Library Preparation: Yulan Piao and Minoru Ko  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC c lone distribution information  
 can be found through the I.M.A.G.E. Consortium/LNLN at:  
 http://image.llnl.gov  
 Plate: NDAM498 row: p column: 01  
 High quality sequence stop: 629.

FEATURES

Location/Qualifiers  
 1..819  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clones="IMAGE:30421320"  
 /tissue\_type="Embryonic Stem cells"  
 /cell\_line="WA01"  
 /lab\_hosts="DH10B (T1 phage-resistant)"  
 /clone\_libs="NIA Human H1 Embryonic Stem Cell cDNA Library (Long)"  
 /notes="Vector: pCMV-Sport6; Site\_1: NotI; Site\_2: SalI;  
 This is a long-transcript enriched cDNA library (Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]) from WA01 cell line. Undifferentiated human ES cell line.WA01/H1 was obtained from WiCell Research Institute, Inc. Madison, WI, cultured according to their instructions, on MEF feeders. They formed round colonies with defined edges and were positive for alkaline phosphatase, SSEA-4, OCT3, OCT4, REX1, UTF, SOX2, CX43 and CX45. They are negative for GATA2, GATA4, PDX1, NCM, MSX1, FLT3, SSEA-1, TUBB3, NES, GFAP, and EOMES. When confluent (18-10 days after plating), the ES cells from 4 X 6cm dishes were treated with 1 mg/ml collagenase, type IV (Invitrogen/GIBCO) for 5-10 min and gently scraped off with 5 ml pipette. RNA was purified with TRIzol Reagent from Invitrogen. Protocol ref: Genome Res. 11: 1553-1558 (2001). [PMID:11544199]) Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen: 5'-PGACTACTTCAGATCGAGCGCCGCTTTTCTTTT-3'] from 3.4g of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to Lona-linker L1-Sal4, purified by phenol/chloroform extraction, and separated from free linkers by Centricon-100 column. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S for 25 cycles. The products were purified by phenol/chloroform extraction and Centricon-100 column. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The average insert size is about 3.6kb."

ORIGIN

Query Match 13.1%; Score 733.4; DB 6; Length 819;  
 Best Local Similarity 94.6%; Pred. No. 1.6e-107;  
 Matches 769; Conservative 0; Mismatches 43; Indels 1; Gaps 1;  
 QY 3891 AGGAGGATACAAAGAGCTAGTCCCAATCGAATCAGAGTATAGAAGATCTTCAATTGA 3950  
 Db 1 AGAGGTATCCCAAGAGCTAGTCCCAATCGAATCAGAGTATAGAAGATCTTCAATTGA 60  
 QY 3951 AAACCATCTAGCTAGCATTTTACTGAGCATTTTAAACTCAGCTTCACAGATGCTTTT 4010  
 Db 61 AAACCATCTAGCTAGCATTTTACTGAGCATTTTAAACTCAGCTTCACAGATGCTTTT 120  
 QY 4011 GTGATGTGATCTTAGCAGTTTGGCCGAGAGAGAAATATCCAGTACCATGCTGTTTT 4070  
 Db 121 GTGATGTGATCTTAGCAGTTTGGCCGAGAGAGAAATATCCAGTACCATGCTGTTTT 180



```

Db      603 ATACTCATTTGGTATATCCAGTAGCTATCTCTCTCAGTTGGTTTGTATGAACAGAGAG 662
Qy      4827 GCCAGCAAC-TTTCTTTGTAAGGCTGGTTAGTAAATATTGCGAGCCACCTG-TGTC 4884
Db      663 GCCAGCAACCTTTCTTTGTAAGGCTGGTTAGTAAATATTGCGAGCCACCTGNTGC 722
Qy      4885 TTGTCATACATCTCTCTG-CGTGTTGTTAGTTGTTGTTTTCATAAACCCCTCTAA 4942
Db      723 TTGCGCATACATCTCTCTGCTGCTGTTAAATTTGTTTTCATAAACCCCTCTAA 782
Qy      4943 AAATGTAACCAACCATGTTTA--GCTTCAGCTGTACAAAA 4981
Db      783 AAATGTAACCAACCATGTTTAAAGCTTGCAGCTGTACAAAA 823

RESULT 14
BG196325
LOCUS      773 bp mRNA linear EST 21-APR-2001
DEFINITION RST15540 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION  BG196325
VERSION     BG196325.1 GI:13718012
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 773)
AUTHORS     Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
            Cain,S., Leventhal,C., Thornton,M., Ramachandran,R.,
            Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Boozer,S.,
            Mays,R., Smith,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K.,
            Offenbacher,J., Danzig,J. and Ducar,M.
            Creation of genome-wide protein expression libraries using random
            activation of gene expression
            Nat. Biotechnol. 19 (5), 440-445 (2001)
JOURNAL     21227151
MEDLINE     11329013
PUBMED
COMMENT     Contact: Scott J. Cain
            Athersys, Inc.
            3201 Carnegie Ave, Cleveland, OH 44115, USA
            Tel: 216 431 9900
            Fax: 216 361 9596
            Email: scain@athersys.com
            High quality sequence stop: 550.
            Location/Qualifiers
FEATURES    source
            1..773
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone_lib="Athersys RAGE Library"
            /note="See 'Creation of Genome-wide Protein Expression
            Libraries using Random Activation of Gene Expression',
            Nature Biotechnology, in press. Note that even though the
            cell type indicated is HT1080, since a random activation
            method was used, these sequence tags are not necessarily
            expressed in HT1080 under normal circumstances."
ORIGIN
Query Match      12.88; Score 715.2; DB 4; Length 773;
Best Local Similarity 97.58; Pred. No. 1.3e-104;
Matches 750; Conservative 0; Mismatches 9; Indels 10; Gaps 2;
Qy      3766 GGAGGGCTTATCGAGCAAAATTGAAGTTTGGACCATTTGTGCTAGCAACATTGATTA 3825
Db      4 GGAGGGCTTATCGAGCAAAATTGAAGTTTGGACCATTTGTGCTAGCAACATTGATTA 63
Qy      3826 TGTTTACTATATCTCATTTTTCCTGAGC-----AGTTATTCACCTTAAGCGGAG 3878
Db      64 TGTTTACTATATCTCATTTTTCCTGAGCAGCATTTAGTTATTCGATTTAGCGGAG 123
Qy      3879 ATATTTCCAGAGGAGGATPACAAAAAGAGCTAGTCCCAATCGAATCAGATAGTAGAAG 3938

```

```

Db      124 ATATTTCCAGAGGAGGATACAAAAGAGCTAGTCCCAATCGAATCAGAGTAGTAGAAG 183
Qy      3939 ATCTTCATTGTAAGAACCACTTACCTCAGCATTTTACTGAGCATTTTAAATCAGCTTCAC 3998
Db      184 ATCTTCATTGTAAGAACCACTTACCTCAGCATTTTACTGAGCATTTTAAATCAGCTTCAC 243
Qy      3999 AGAGATGTCTTTGTGATGTGATCTAGCAGTTTGGCCGAGGAGAAATATCCAGTA 4058
Db      244 AGAGATGTCTTTGTGATGTGATCTAGCAGTTTGGCCGAGGAGAAATATCCAGTA 303
Qy      4059 CCATGCTGTTTGTGGCATGAATATAGCCCACTGACCCAGGAAATATTAAACCAACCCACT 4118
Db      304 CCATGCTGTTTGTGGCATGAATATAGCCCACTGACCCAGGAAATATTAAACCAACCCACT 363
Qy      4119 GAAACACTTGTGTGCGAGCAGCTCTGAACCTGATTTTAAAGAAATTTGCTCATGGA 4178
Db      364 GAAACACTTGTGTGTGAGCAGCTCTGAACCTGATTTTAAAGAAATTTGCTCATGGA 423
Qy      4179 CTTGTCATCTCTTTTATAAAGGCTCAGTCAGAGGAGACAGCTGTTAAATTTCCACAGC 4238
Db      424 CTTGTCATCTCTTTTATAAAGGCTCAGTCAGAGGAGACAGCTGTTAAATTTCCACAGC 483
Qy      4239 AATCATTCAGACTAACTTTTATTAGGAGAGCCCTATGCCAGCTGGAGTATGCTAAGA 4298
Db      484 AATCATTCAGACTAACTTTTATTAGGAGAGCCCTATGCCAGCTGGAGTATGCTAAGA 543
Qy      4299 GGCTCCAGCTTTGCATTCGAAAGCTTTTGTCTAAAGTTTGCAC---TTTTCCTTTTTC 4355
Db      544 GGCTCCAGCTTTGCATTCGAAAGCTTTTGTCTAAAGTTTGCACCTTTTTCCTTTTTC 603
Qy      4356 ATTTCCTATTTTAAAGTAGTTTACTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 4415
Db      604 ATTTCCTATTTTAAAGTAGTTTACTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 663
Qy      4416 ATTGGATGCTAAACCTTTTATTAGAGTATTTTAAATATGAGCAATATCACCCTC 4475
Db      664 ATGGGATGCTAAACCTTTTATTAGAGTATTTTAAATATGAGCAATATCACCCTC 723
Qy      4476 TTATTGACATACCTAAATATGAGTATTTTAAATATGAGCAATATCACCCTC 4524
Db      724 TTATTGACATACCTAAATATGAGTATTTTAAATATGAGCAATATCACCCTC 772

RESULT 15
BG194321
LOCUS      795 bp mRNA linear EST 21-APR-2001
DEFINITION RST13466 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION  BG194321
VERSION     BG194321.1 GI:13716008
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 795)
AUTHORS     Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
            Cain,S., Leventhal,C., Thornton,M., Ramachandran,R.,
            Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Boozer,S.,
            Mays,R., Smith,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K.,
            Offenbacher,J., Danzig,J. and Ducar,M.
            Creation of genome-wide protein expression libraries using random
            activation of gene expression
            Nat. Biotechnol. 19 (5), 440-445 (2001)
JOURNAL     21227151
MEDLINE     11329013
PUBMED
COMMENT     Contact: Scott J. Cain
            Athersys, Inc.
            3201 Carnegie Ave, Cleveland, OH 44115, USA
            Tel: 216 431 9900
            Fax: 216 361 9596
            Email: scain@athersys.com
            High quality sequence stop: 441.

```

FEATURES  
source

## Location/Qualifiers

1..795  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/cell\_line="HT1080"  
/clone\_lib="Athersys RAGE Library"  
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', the Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

## ORIGIN

Query Match 12.7%; Score 710.8; DB 4; Length 795;  
Best Local Similarity 97.1%; Pred. No. 6.8e-104;  
Matches 775; Conservative 0; Mismatches 16; Indels 7; Gaps 5;

QY 3765 TGGAGGCTTATCAGACCAATTGAAGTTTGGACCCATTGTGTACTAGCAACATTGATT 3824  
Db 4 TGGAGGACTATCAGACCAATTG-AGTTTGGACCCATTGTGTACTAGCAACATTGATT 62

QY 3825 ATGTTTACTATATTCTCATTTTTCCTGAGCAGTTAATTGCACCTTAAGCGGAAGATATT 3884  
Db 63 ATGTTTACTATATTCTCATTTTTCCTGAGCAGTTAATTGCACCTTAAGCGGAAGATATT 122

QY 3885 CCCAGAGGAGGATACAAAGAAGCTAGTCCCAATCGATCAGATATAGAAGATCTTC 3944  
Db 123 CCCTGAAGAGGATACAAAGAAGCTAGTCCCAATCGATCAGATATAGAAGATCTTC 182

QY 3945 ATTTGAAACCATTCTACCTCAGCATTTACTGAGCATTTTAAACTCAGCTTCACAGAT 4004  
Db 183 ATTTGAAACCATTCTACCTCAGCATTTACTGAGCATTTTAAACTCAGCTTCACAGAT 242

QY 4005 GTCCTTTGTGATGTCATGCTTAGCAGTTTGGCCCAAGAGAAATATCCAGTACCATGC 4064  
Db 243 GTCCTTTGTGATGTCATGCTTAGCAGTTTGGCCCAAGAGAAATAT-CAGTACCATGC 301

QY 4065 TGTTTGTGGCATGAATATAGCCACCTGACTAGGAATTTTACCAACCCACTGAAAC 4124  
Db 302 TGTTTGTGGCATGAATATAGCCACCTGACTAGGAATTTTACCAACCCACTGAAAC 361

QY 4125 TTGTGTGCGAGCAGCTCTGAACCTGATTTTACTTTTAAAGAAATTTGCTATGGACCTGTC 4184  
Db 362 TTGTGTGTTGAGCAGCTCTGAACCTGATTTTACTTTTAAAGAAATTTGCTATGGACCTGTC 421

QY 4185 ATCCTTTTATAAAGCTCACTGACAGAGACAGCTGTTAATTTCCACAGCAATCAT 4244  
Db 422 ATCCTTTTATAAAGCTCACTGACAGAGACAGCTGTTAATTTCCACAGCAATCAT 481

QY 4245 TGCAGACTAATTTTATAGGAGAGCCCTATGCCAGCTGGGAGTGAATGCTAAGAGGCTCC 4304  
Db 482 TGCAGACTAATTTTATAGGAGAGCCCTATGCCAGCTGGGAGTGAATGCTAAGAGGCTCC 541

QY 4305 AGCTTTTCATTCCAAAGCCCTTTGCTAAAGTTTGGCACTTTTTCATTTCCCAT 4364  
Db 542 AGCTTTTCATTCCAAAGCCCTTTGCTAAAGTTTGGCACTTTTTCATTTCCCAT 598

QY 4365 TTTTAAAGTAGTACTAAGTAACTAGTATTTCTTCTCTCTGAGTATACGAATTTGGGATG 4424  
Db 599 TTTTAAAGTAGTACTAAGTAACTAGTATTTCTTCTCTCTGAGTATACGAATTTGGGATG 658

QY 4425 TCTAAACCTATTTTATAGATGTTATTTAAATATAGCAGCAATATCACCTCTTATTCACA 4484  
Db 659 TCTAAACCTA-TTTTATAGATGTTATTTAAATATAGCAGCAATATCACCTCTTATTCANC 717

QY 4485 ATACCTAAATATAGATTTTATTAATTTAAGACTGTAATGCTTTAACCCTAACT 4544  
Db 718 ATACCTAAATATAGATTTTATTAATTTAAGACTGTAATGCTTTAACCCTAACT 776

QY 4545 ACTGAAGAGCTCAATGAT 4562

Db 777 ACTGAAGAGCTCAATGAT 794

Search completed: November 22, 2004, 10:37:19  
Job time : 15717 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2004, 12:38:18 ; Search time 106.597 Seconds  
(without alignments)  
3122.980 Million cell updates/sec

Title: US-10-023-888-4

Perfect score: 4907  
Sequence: 1 MLFKLQRQRYTCLSHRYGL.....SLKTQLAYFTDKNTRQLX 928

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729293 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : A\_Geneseq\_23Sep04:\*

- 1: geneseq1980s:\*
- 2: geneseq1990s:\*
- 3: geneseq2000s:\*
- 4: geneseq2001s:\*
- 5: geneseq2002s:\*
- 6: geneseq2003as:\*
- 7: geneseq2003bs:\*
- 8: geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4907	100.0	928	4	ABR61376 Human Glc
2	4907	100.0	928	7	ADD27812 Human Glc
3	4907	100.0	928	7	ABW01488 Human Glc
4	4907	100.0	928	7	ABW01537 Human Glc
5	4896	99.8	1256	5	Aae25290 Human nuc
6	4830	94.4	1199	7	ADD27810 Soluble h
7	4830	94.4	1199	7	ABW01487 N-acetylgl
8	4630	94.4	1199	7	ABW01536 N-acetylgl
9	4500	91.7	1196	5	Aae25294 Human nuc
10	4468	91.1	846	6	ADA54824 Human pro
11	4461	90.9	847	7	ADG39856 Protein s
12	4391	89.5	1459	6	ABU07381 Human pro
13	4391	89.5	1459	7	ADG39794 Human nov
14	3894	79.4	908	4	ABR61384 Murine gl
15	3894	79.4	908	7	ADD27817 GLCNAC-ph
16	3894	79.4	908	7	ABW01491 Mouse pro
17	3894	79.4	908	7	ABW01540 Mouse pro
18	2016	41.1	663	7	ADG39855 Protein s
19	1900	38.7	367	4	AAW16457 Peptide #
20	1900	38.7	367	4	AAW28952 Peptide #
21	1900	38.7	367	4	ABE30279 Peptide #
22	1900	38.7	367	4	AAW56268 Human bra
23	1900	38.7	367	4	AAW04186 Peptide #
24	1900	38.7	367	5	ABG38224 Human pep
25	1900	38.7	367	8	ABO59293 Human gen

26	1286	26.2	242	7	ADD26807 Human adi
27	923	18.8	178	6	ABU11844 Human MDJ
28	734.5	15.0	384	7	ADG39857 Protein s
29	503	10.3	113	4	ABR61382 Partial r
30	503	10.3	113	7	ADD27822 Rat alpha
31	503	10.3	113	7	ABW01494 Partial r
32	503	10.3	113	7	ABW01543 Partial r
33	488	9.9	651	7	ADG39859 Protein s
34	488	9.9	652	4	ABE59094 Drosophila
35	479	9.8	502	4	ABR61383 Partial D
36	479	9.8	502	7	ADD27824 Fruit fly
37	479	9.8	502	7	ABW01495 Partial f
38	479	9.8	502	7	ABW01544 Partial f
39	380	7.7	67	8	ABO59064 Human gen
40	360	7.3	71	7	ADL21059 Novel num
41	349	7.1	68	4	ABB38187 Peptide #
42	349	7.1	68	4	AAW31616 Peptide #
43	349	7.1	68	4	AAW71337 Human bon
44	349	7.1	68	4	AAW58821 Human bra
45	349	7.1	68	4	ABG53045 Human liv

## ALIGNMENTS

## RESULT 1

ABR61376  
ID ABR61376 standard; protein; 928 AA.

XX ABR61376;

XX AC

XX 01-AUG-2003 (first entry)

XX Human GlcNAc-phosphotransferase alpha-subunit.

XX Human; N-acetylglucosamine-1-phosphotransferase; nephrotropic;

KW GlcNAc-phosphotransferase; phosphodiester alpha-GlcNAcase;

KW N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase;

KW enzyme replacement therapy; phosphorylated lysosomal hydrolase;

KW lysosomal storage disease; enzyme; alpha-subunit.

XX Homo sapiens.

XX US6537785-B1.

XX 25-MAR-2003.

XX 10-AUG-2000; 2000US-00635077.

XX 14-SEP-1999; 99US-0153831P.

XX (GENZ-) GENZYME GLYCOBIOLOGY RES INST INC.

XX Canfield WM;

XX WFI; 2001-290356/30.

XX N-PSDB; ACC81001.

XX Novel N-acetylglucosamine-1-phosphotransferase and N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase, useful for producing phosphorylated lysosomal hydrolase for treating lysosomal storage diseases.

XX Disclosure; Page 26-28; 62pp; English.

XX The invention relates to a novel isolated human N-acetylglucosamine-1-phosphotransferase (GlcNAc-phosphotransferase) (I) and phosphodiester alpha-GlcNAcase (N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase) (II). The protein of the invention has nephrotropic activity, and may be useful in enzyme replacement therapy. A protein of the invention (I), (II) is useful for preparing a phosphorylated lysosomal hydrolase. The phosphorylated hydrolase comprising a terminal mannose-6-phosphate, is useful for treating a

CC patient suffering from a lysosomal storage disease. The present sequence  
 CC is used in the exemplification of the invention

XX SQ Sequence 928 AA;

	Query Match	Best Local Similarity	100.0%; Score 4907; DB 4; Length 928;	Mismatches	0; Indels	0; Gaps	0;
Qy	1	MLFKLLQRTQTYTCLSHRYGLVYVCFGLGVVVTIVSAFQGEVWLEWSDQYHVLFDSDYRDN	60				
Db	1	MLFKLLQRTQTYTCLSHRYGLVYVCFGLGVVVTIVSAFQGEVWLEWSDQYHVLFDSDYRDN	60				
Qy	61	AGKSFQNRCLCLPMPIDVVYTWNGTDLLELLKELQVREQMEEEKAMREILGKNTTEPTK	120				
Db	61	AGKSFQNRCLCLPMPIDVVYTWNGTDLLELLKELQVREQMEEEKAMREILGKNTTEPTK	120				
Qy	121	KSEKQLECLLTHCIVKPNMLVLDPALPANITLKVPSLYPSFHSASDIFNVAKPNPSTNV	180				
Db	121	KSEKQLECLLTHCIVKPNMLVLDPALPANITLKVPSLYPSFHSASDIFNVAKPNPSTNV	180				
Qy	181	SVVVFDSKQVEDAHSGLLKGNRSQTVWRGYLTDDKEVPGVLMDLAFSLGPPPTFKET	240				
Db	181	SVVVFDSKQVEDAHSGLLKGNRSQTVWRGYLTDDKEVPGVLMDLAFSLGPPPTFKET	240				
Qy	241	NQTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKQNTIDGKELTISPA	300				
Db	241	NQTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKQNTIDGKELTISPA	300				
Qy	301	YLLWDLAISOSKODEDISASRFEDNEELYSLSIERHAPWVRNIFIVTNGQIPSWLNL	360				
Db	301	YLLWDLAISOSKODEDISASRFEDNEELYSLSIERHAPWVRNIFIVTNGQIPSWLNL	360				
Qy	361	DNPRVTIVTHQDVPRNLSHLPTFFSSPAIESHIHRIEGLSKFIYLDNDVDMFGKVPDDF	420				
Db	361	DNPRVTIVTHQDVPRNLSHLPTFFSSPAIESHIHRIEGLSKFIYLDNDVDMFGKVPDDF	420				
Qy	421	YSHSGKQVILTVPNVPCAGCPGSIWKDGYCDKACNNSACDWDGDCSNGSGSYIAG	480				
Db	421	YSHSGKQVILTVPNVPCAGCPGSIWKDGYCDKACNNSACDWDGDCSNGSGSYIAG	480				
Qy	481	GGGTGSGVGHWPQFQGGINSVSCNCGCANSMLADKFCDOACNVLSCGFDAGDCQDHF	540				
Db	481	GGGTGSGVGHWPQFQGGINSVSCNCGCANSMLADKFCDOACNVLSCGFDAGDCQDHF	540				
Qy	541	HELYKVILLPNQTHYIIPKGECLPYFSPAFAVAKRGVEGAYSDNPIIRHASIANKWTIHL	600				
Db	541	HELYKVILLPNQTHYIIPKGECLPYFSPAFAVAKRGVEGAYSDNPIIRHASIANKWTIHL	600				
Qy	601	IMHSGMATTIHENLTQNTNDEEFKMOITVEVDTRGPKLNSTAGKYNLSPITLLP	660				
Db	601	IMHSGMATTIHENLTQNTNDEEFKMOITVEVDTRGPKLNSTAGKYNLSPITLLP	660				
Qy	661	EAEILFEDIPEKKEFPFKHSDVNSTRAQEEVKIPLVNIISLLPKDAQLSLNTLDLQLEH	720				
Db	661	EAEILFEDIPEKKEFPFKHSDVNSTRAQEEVKIPLVNIISLLPKDAQLSLNTLDLQLEH	720				
Qy	721	GDITLKGYNISKSALLSFLMNSCHAKIKNOAIIITDNDLSVAPQEKQVHKSILPNSLG	780				
Db	721	GDITLKGYNISKSALLSFLMNSCHAKIKNOAIIITDNDLSVAPQEKQVHKSILPNSLG	780				
Qy	781	VSELRQLATTPAVSVKYNHGDQGNPPDLLETTARFRVETHQTKTIGGNVTEKPPSLIV	840				
Db	781	VSELRQLATTPAVSVKYNHGDQGNPPDLLETTARFRVETHQTKTIGGNVTEKPPSLIV	840				
Qy	841	PLESQMTKEKIKTCKEKENMERMEANAENHIGVTEVLLGRKLQHYTDSYLGFLPWEKKYF	900				
Db	841	PLESQMTKEKIKTCKEKENMERMEANAENHIGVTEVLLGRKLQHYTDSYLGFLPWEKKYF	900				
Qy	901	QDLIDDEESLKTOLAYFTDSKNTGRQLK	928				
Db	901	QDLIDDEESLKTOLAYFTDSKNTGRQLK	928				

RESULT 2

ADD27812

ID ADD27812 standard; protein; 928 AA.

XX AC ADD27812;

XX DT 15-JAN-2004 (first entry)

XX DE Human GlcNAc-phosphotransferase alpha subunit.

XX KW human; protein phosphorylation; soluble GlcNAc-phosphotransferase;

XX KW UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease.

XX OS Homo sapiens.

XX PN US2003119088-A1.

XX PD 26-JUN-2003.

XX PF 21-DEC-2001; 2001US-00023888.

XX PR 21-DEC-2001; 2001US-00023888.

XX PA (NOVA-) NOVATIME PHARM INC.

XX PI Canfield W, Kudo M;

XX DR WPI; 2003-801323/75.

XX N-PSDB; ADD27811.

PT Phosphorylating a protein for treating a patient suffering from a

PT lysosomal storage disease e.g. Fabry's disease by contacting the protein

PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated

PT protein.

XX PS Claim 7; SEQ ID NO 4; 55pp; English.

XX CC The invention relates to a method of phosphorylating a protein comprising

XX CC contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-

XX CC acetylglucosamine) and producing a phosphorylated protein. The method is

XX CC useful for treating a patient suffering from a lysosomal storage disease

XX CC e.g. Fabry's disease. The present sequence represents the amino acid

XX CC sequence of the human GlcNAc-phosphotransferase alpha subunit.

XX SQ Sequence 928 AA;

Query Match 100.0%; Score 4907; DB 7; Length 928;

Best Local Similarity 100.0%; Pred. No. 0;

Mismatches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLFKLLQRTQTYTCLSHRYGLVYVCFGLGVVVTIVSAFQGEVWLEWSDQYHVLFDSDYRDN

Db 1 MLFKLLQRTQTYTCLSHRYGLVYVCFGLGVVVTIVSAFQGEVWLEWSDQYHVLFDSDYRDN

Qy 61 AGKSFQNRCLCLPMPIDVVYTWNGTDLLELLKELQVREQMEEEKAMREILGKNTTEPTK

Db 61 AGKSFQNRCLCLPMPIDVVYTWNGTDLLELLKELQVREQMEEEKAMREILGKNTTEPTK

Qy 121 KSEKQLECLLTHCIVKPNMLVLDPALPANITLKVPSLYPSFHSASDIFNVAKPNPSTNV

Db 121 KSEKQLECLLTHCIVKPNMLVLDPALPANITLKVPSLYPSFHSASDIFNVAKPNPSTNV

Qy 181 SVVVFDSKQVEDAHSGLLKGNRSQTVWRGYLTDDKEVPGVLMDLAFSLGPPPTFKET

Db 181 SVVVFDSKQVEDAHSGLLKGNRSQTVWRGYLTDDKEVPGVLMDLAFSLGPPPTFKET

Qy 241 NQTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKQNTIDGKELTISPA

Db 241 NQTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKQNTIDGKELTISPA

Qy 301 YLLWDLAISOSKODEDISASRFEDNEELYSLSIERHAPWVRNIFIVTNGQIPSWLNL

Db 301 YLLWDLAISOSKODEDISASRFEDNEELYSLSIERHAPWVRNIFIVTNGQIPSWLNL



Db 661 EAEILFEDIPKEKPPKFRHVNSTRRAQAEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720  
 QY 721 GDITLKGYNLSKALLRSFLMNSOHAKIKNOAIITDETNDLSLVAPOKQVHKSLPNSLG 780  
 Db 721 GDITLKGYNLSKALLRSFLMNSOHAKIKNOAIITDETNDLSLVAPOKQVHKSLPNSLG 780  
 QY 781 VSELRQRLTTPAVSVKVNHGDPQGNPPDLDTTARFRVETHQKIGNTVKEKPPSLIV 840  
 Db 781 VSELRQRLTTPAVSVKVNHGDPQGNPPDLDTTARFRVETHQKIGNTVKEKPPSLIV 840  
 QY 841 PLESQMTKEKITGKEKENSMEENAHGVTGLGRKLOHYTDSYLGFLPWEKKYF 900  
 Db 841 PLESQMTKEKITGKEKENSMEENAHGVTGLGRKLOHYTDSYLGFLPWEKKYF 900  
 QY 901 QDLDEESLTKQLAYFTDSKNTGRQLK 928  
 Db 901 QDLDEESLTKQLAYFTDSKNTGRQLK 928

## RESULT 4

ABW01537 standard; protein; 928 AA.

ID ABW01537 standard; protein; 928 AA.

XX AC ABW01537;

XX DT 15-JAN-2004 (first entry)

XX DE Human GlcNAc-6-phosphotransferase alpha subunit precursor protein.

XX KW Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;

XX KW N-acetylglucosamine-1-phosphotransferase; enzyme; gene therapy; human.

XX OS Homo sapiens.

XX FN US2003124653-A1.

XX PD 03-JUL-2003.

XX PF 21-DEC-2001; 2001US-00023890.

XX PR 21-DEC-2001; 2001US-00023890.

XX PA (NOVA-) NOVAZYME PHARM INC.

XX PI Canfield WM;

XX DR WPI; 2003-810985/76.

XX DR N-FSDB; AAD62650.

Producing a glycoprotein with reduced complex carbohydrates by culturing the lectin resistant mammalian cell expressing the glycoprotein for treating lysosomal storage disease.

XX Claim 10; Page 16-18; 46pp; English.

XX CC The present invention provides a method of producing a glycoprotein

having reduced complex carbohydrates by culturing the lectin resistant mammalian cell expressing the glycoprotein. The method is useful for producing a glycoprotein with reduced complex carbohydrates for treating lysosomal storage disease. The present invention is also useful in gene therapy. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-6-phosphotransferase alpha subunit precursor protein

XX SQ Sequence 928 AA;

XX Query Match 100.0%; Score 4907; DB 7; Length 928;

XX Best Local Similarity 100.0%; Pred. No. 0;

XX Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFKLQRYTYTCLSHRYGLYVCFGLGVVTVISAFQGEVLEWSDQVHVLFDSDYRNI 60

Db 1 MLFKLQRYTYTCLSHRYGLYVCFGLGVVTVISAFQGEVLEWSDQVHVLFDSDYRNI 60

QY 61 AGKSFQNRCLPMPIDVVYTWNGTDLLELLKELQVREQMEEEQKAMEIILGKNTTEPTK 120  
 Db 61 AGKSFQNRCLPMPIDVVYTWNGTDLLELLKELQVREQMEEEQKAMEIILGKNTTEPTK 120  
 QY 121 KSEKQLECLLTHICIKVPMVLDPALPANITLKVPSLPSFHSASDIPNVAKPKNPSNV 180  
 Db 121 KSEKQLECLLTHICIKVPMVLDPALPANITLKVPSLPSFHSASDIPNVAKPKNPSNV 180  
 QY 181 SVVVFDSKDVDEASHGLLKGNRSOTVMRGYLTITDKVEPGLVLMODLAFISGFPTPKET 240  
 Db 181 SVVVFDSKDVDEASHGLLKGNRSOTVMRGYLTITDKVEPGLVLMODLAFISGFPTPKET 240  
 QY 241 NOLKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFOELNKQTKNMTIDGKELTISPA 300  
 Db 241 NOLKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFOELNKQTKNMTIDGKELTISPA 300  
 QY 301 YLLWDLISAISSQKODEDISASRFEDNEELRYSLRSIERHAPWVRNIFIVTNGQIPSWLNL 360  
 Db 301 YLLWDLISAISSQKODEDISASRFEDNEELRYSLRSIERHAPWVRNIFIVTNGQIPSWLNL 360  
 QY 361 DNPRTVITHQDVFRNLSHLPTFFSPATIESHIHRIEGLSQKFIYLNDDVMFGKDWPPDF 420  
 Db 361 DNPRTVITHQDVFRNLSHLPTFFSPATIESHIHRIEGLSQKFIYLNDDVMFGKDWPPDF 420  
 QY 421 YSHSKGQKYLTPVPCNCAEGCPGSKIWDGYCDKACNNSACDWDGDCSGNSGSRITAG 480  
 Db 421 YSHSKGQKYLTPVPCNCAEGCPGSKIWDGYCDKACNNSACDWDGDCSGNSGSRITAG 480  
 QY 481 GGGTSGISGVHPWFGGGINSVSYCNQGCANSLADKFCDOACNVLSGCGFDAGCGDHF 540  
 Db 481 GGGTSGISGVHPWFGGGINSVSYCNQGCANSLADKFCDOACNVLSGCGFDAGCGDHF 540  
 QY 541 HELYKVILLPNQTHYIIPKGECLPYFSPAEVAKRGVGAISDNPIIRHASTANKWKTILH 600  
 Db 541 HELYKVILLPNQTHYIIPKGECLPYFSPAEVAKRGVGAISDNPIIRHASTANKWKTILH 600  
 QY 601 IMHSGMNTTTHFNLTFTONTNDEBEFKQITVEVDTRGPKLNSTAKGYENLVSPITLLP 660  
 Db 601 IMHSGMNTTTHFNLTFTONTNDEBEFKQITVEVDTRGPKLNSTAKGYENLVSPITLLP 660  
 QY 661 EAEILFEDIPKEKPPKFRHVNSTRRAQAEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720  
 Db 661 EAEILFEDIPKEKPPKFRHVNSTRRAQAEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720  
 QY 721 GDITLKGYNLSKALLRSFLMNSOHAKIKNOAIITDETNDLSLVAPOKQVHKSLPNSLG 780  
 Db 721 GDITLKGYNLSKALLRSFLMNSOHAKIKNOAIITDETNDLSLVAPOKQVHKSLPNSLG 780  
 QY 781 VSELRQRLTTPAVSVKVNHGDPQGNPPDLDTTARFRVETHQKIGNTVKEKPPSLIV 840  
 Db 781 VSELRQRLTTPAVSVKVNHGDPQGNPPDLDTTARFRVETHQKIGNTVKEKPPSLIV 840  
 QY 841 PLESQMTKEKITGKEKENSMEENAHGVTGLGRKLOHYTDSYLGFLPWEKKYF 900  
 Db 841 PLESQMTKEKITGKEKENSMEENAHGVTGLGRKLOHYTDSYLGFLPWEKKYF 900  
 QY 901 QDLDEESLTKQLAYFTDSKNTGRQLK 928  
 Db 901 QDLDEESLTKQLAYFTDSKNTGRQLK 928

## RESULT 5

AAE25290

ID AAE25290 standard; protein; 1256 AA.

XX AC AAE25290;

XX DT 30-OCT-2002 (first entry)

XX DE Human nucleic acid-associated protein (NAAP-9).

XX KW Human; nucleic acid-associated protein; NAAP-9; neurological disorder; arteriosclerosis; cancer; cell proliferative disorder; atherosclerosis;

KW lymphoma; epilepsy; Alzheimer's disease; developmental; anticonvulsant;  
KW autoimmune disorder; AIDS; allergy; anaemia; stroke; malaria; leishmania;  
KW gene therapy; nontropic; neuroprotective; cerebroprotective; virucide;  
KW immunosuppressive; protozoicide; antimicrobial.

OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Domain 448..469 /note= "Notch domain"  
FT Domain 500..536 /note= "Notch domain"  
FT Domain 1018..1030 /note= "EF-hand calcium-binding domain"

PN WO200250279-A2.  
XX 27-JUN-2002.  
PD 19-DEC-2001; 2001WO-US050256.  
XX 21-DEC-2000; 2000US-0257714P.  
PR 05-JAN-2001; 2001US-0260081P.  
PR 16-JAN-2001; 2001US-0262302P.  
PR 23-JAN-2001; 2001US-0263823P.  
PR 02-FEB-2001; 2001US-0266088P.  
PR 29-OCT-2001; 2001US-0348442P.

XX (INCY-) INCYTE GENOMICS INC.  
XX Baughin MR, Lu Y, Arvizu C, Ramkumar J, Yao MG, Policky JL;  
PI Walla NK, Tribouley KM, Yue H, Batra S, Ding L, Lal PG;  
PI Borowsky M, Lu DAM, Gandhi AR, Griffin JA, Xu Y, Azimzai Y;  
PI Gietzen KJ, Tang YT, Warren BA, Mason PM, Burford N, Hafalia AJA;  
PI Lee EA, Yang J, Gorvad AE, Emerling BM, Marquis JP, Lee SY;  
PI Swarnakar A, Reddy R;  
XX WPI; 2002-519887/55.  
DR N-P5DB; RAD41199.

XX Nucleic acid associated proteins and nucleic acids for diagnosing,  
PT treating and preventing cell proliferative (e.g. cancers), neurological  
XX (e.g. epilepsy or stroke) or autoimmune disorders (e.g. AIDS).  
PS Claim 64; Page 163-165; 193pp; English.  
XX The invention relates to nucleic acid-associated proteins (NAAP) and  
CC nucleic acids. The nucleic acid and amino acid sequences are useful for  
CC diagnosing, treating and preventing cell proliferative e.g.  
CC arteriosclerosis, atherosclerosis, lymphoma or cancers), neurological  
CC (e.g. epilepsy, Alzheimer's disease or stroke), developmental, and  
CC autoimmune disorders (e.g. AIDS, allergies, or anaemia) or infections  
CC (e.g. malaria, or leishmania), as well as in assessing the effects of  
CC exogenous compound on the expression of nucleic acid and amino acid  
CC sequences of nucleic acid-associated proteins. The invention is useful in  
CC gene therapy. The present sequence is human NAAP-9

SQ Sequence 1256 AA;  
Query Match 99.8%; Score 4896; DB 5; Length 1256;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 926; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLFKLQRTYTCLSHRYGLVCFGLGVVTVTSAPFGVLEWSRDQYHVLFDSDNI 60  
DB 1 MLFKLQRTYTCLSHRYGLVCFGLGVVTVTSAPFGVLEWSRDQYHVLFDSDNI 60  
QY 61 AGKSFQNRCLCLPMPIDVYVTVWNGTDLELLKQVREQMEEQKAMREILGKNTTEPTK 120  
DB 61 AGKSFQNRCLCLPMPIDVYVTVWNGTDLELLKQVREQMEEQKAMREILGKNTTEPTK 120  
QY 121 KSEKQLECLLTCHIKVPMVLDPALPANITLKDPSLYPSFHSASDIFNVAKPKNPSTNV 180

DB 121 KSEKQLECLLTCHIKVPMVLDPALPANITLKDPSLYPSFHSASDIFNVAKPKNPSTNV 180  
QY 181 SVVVFSTKDVDEAHSGLLKGNRSQTVWRYGLTTTKEVGLVLMODLAFSLGPPFPFKET 240  
DB 181 SVVVFSTKDVDEAHSGLLKGNRSQTVWRYGLTTTKEVGLVLMODLAFSLGPPFPFKET 240  
QY 241 NQLTKLPENLSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNMTIDGKELTISPA 300  
DB 241 NQLTKLPENLSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNMTIDGKELTISPA 300  
QY 301 YLLWDLISAISQKQDEDI SASRPEDNEELRYSLRSTIERHAPVWRNIFIVTNGQISFWNL 360  
DB 301 YLLWDLISAISQKQDEDI SASRPEDNEELRYSLRSTIERHAPVWRNIFIVTNGQISFWNL 360  
QY 361 DNPRTIVTHQDVFRNLHLPTFFSSPAIESHRIHRIEGLSQFIYLLNDDVFMFGKVPDPDF 420  
DB 361 DNPRTIVTHQDVFRNLHLPTFFSSPAIESHRIHRIEGLSQFIYLLNDDVFMFGKVPDPDF 420  
QY 421 YSHSGKQKYLTPVPNCAGCGPGSWIKDGYCDKACNNSACDWDGDCSGNSGGSYIAG 480  
DB 421 YSHSGKQKYLTPVPNCAGCGPGSWIKDGYCDKACNNSACDWDGDCSGNSGGSYIAG 480  
QY 481 GGGTSGIGVGHPMQFGGINSVSYCNQGCANSWLADKFCDOACNVLSGCFDAGDCQDHF 540  
DB 481 GGGTSGIGVGHPMQFGGINSVSYCNQGCANSWLADKFCDOACNVLSGCFDAGDCQDHF 540  
QY 541 HELYKVILLPNQTHYIIPKGECLPYPSFAEAKRGVEGAYSDNPIIRHASIANKWTIHL 600  
DB 541 HELYKVILLPNQTHYIIPKGECLPYPSFAEAKRGVEGAYSDNPIIRHASIANKWTIHL 600  
QY 601 IMHSGNNTIHNLTFTQNTDEEFKQITVEVDTRGPKLNSTAGKGYENLVSPITLLP 660  
DB 601 IMHSGNNTIHNLTFTQNTDEEFKQITVEVDTRGPKLNSTAGKGYENLVSPITLLP 660  
QY 661 EABILFEDIPKEKRPFKRHDVNSTRAQEEVKIPLVNI SLLPKDAQLSLNTLDLQLEH 720  
DB 661 EABILFEDIPKEKRPFKRHDVNSTRAQEEVKIPLVNI SLLPKDAQLSLNTLDLQLEH 720  
QY 721 GDITLKGYNLSKALLRSFLMNSQAKIKNOAIITDETNDLSLVAPOEKVHKSILPNSLG 780  
DB 721 GDITLKGYNLSKALLRSFLMNSQAKIKNOAIITDETNDLSLVAPOEKVHKSILPNSLG 780  
QY 781 VSERLQRLTFPAPVSVKVGHDQGNPDLDTTAFRVETHTKTIGGNVTKEKPSLIIV 840  
DB 781 VSERLQRLTFPAPVSVKVGHDQGNPDLDTTAFRVETHTKTIGGNVTKEKPSLIIV 840  
QY 841 PLESQMTKEKKTITGKEKENSMEENAEHIGVTEVLLGKLOHYTDSYLGFLPWEKKYF 900  
DB 841 PLESQMTKEKKTITGKEKENSMEENAEHIGVTEVLLGKLOHYTDSYLGFLPWEKKYF 900  
QY 901 QDILLDEESLKTQLAYFTDSKNTGROLK 928  
DB 901 QDILLDEESLKTQLAYFTDSKNTGROLK 928

RESULT 6  
ADD27810  
ID ADD27810 standard; protein; 1199 AA.  
XX AC ADD27810;  
XX DT 15-JAN-2004 (first entry)  
XX DE Soluble human GlcNAc-phosphotransferase.  
KW human; protein phosphorylation; soluble GlcNAc-phosphotransferase;  
KW UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease;  
KW enzyme.  
XX OS Synthetic.  
XX OS Homo sapiens.  
XX PN US2003119088-A1.

XX 26-JUN-2003.  
 XX 21-DEC-2001; 2001US-00023888.  
 XX 21-DEC-2001; 2001US-00023888.  
 XX (NOVA-) NOVAZYME PHARM INC.  
 XX Canfield W, Kudo M;  
 XX WPI; 2003-801323/75.  
 XX N-PSDB; ADD27809.  
 XX  
 XX Phosphorylating a protein for treating a patient suffering from a  
 PT lysosomal storage disease e.g. Fabry's disease by contacting the protein  
 PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated  
 PT protein.  
 XX  
 XX Claim 3; SEQ ID NO 2; 55pp; English.  
 XX  
 XX The invention relates to a method of phosphorylating a protein comprising  
 CC contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-  
 CC acetylglucosamine) and producing a phosphorylated protein. The method is  
 CC useful for treating a patient suffering from a lysosomal storage disease  
 CC e.g. Fabry's disease. The present sequence represents the amino acid  
 CC sequence of soluble human GlcNAc-phosphotransferase.  
 XX  
 XX Sequence 1199 AA;  
 SQ

Query Match 94.4%; Score 4630; DB 7; Length 1199;  
 Best Local Similarity 99.3%; Pred. No. 0;  
 Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 45 SRDQHVLFSDYRNIAKSPQNLCLPMPIDVVYTWNGTDLLELLKELQVREQMEEQ 104  
 DB 35 SRDQHVLFSDYRNIAKSPQNLCLPMPIDVVYTWNGTDLLELLKELQVREQMEEQ 94  
 QY 105 KAMREILGKNTTEPTKSEKOLECILLTHCIKVPMLVLDPALPANITLKDVPSPYPSFSA 164  
 DB 95 KAMREILGKNTTEPTKSEKOLECILLTHCIKVPMLVLDPALPANITLKDVPSPYPSFSA 154  
 QY 165 SDIFNVAKPNPSTNVSVVWFDS\*KDVEDAHSGLLKGNRSQTVWRGYLTDDKEVPGVLVM 224  
 DB 155 SDIFNVAKPNPSTNVSVVWFDS\*KDVEDAHSGLLKGNRSQTVWRGYLTDDKEVPGVLVM 214  
 QY 225 QDLAFLSGFPPTFKETVQLTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKOT 284  
 DB 215 QDLAFLSGFPPTFKETVQLTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKOT 274  
 QY 285 KKNMTIDGKELTISPAYLLWDLAISQSKODEDISASRFEDNEELRYSLRSIERHAPWVR 344  
 DB 275 KKNMTIDGKELTISPAYLLWDLAISQSKODEDISASRFEDNEELRYSLRSIERHAPWVR 334  
 QY 345 NIFVTNGQIPSWNLNLPRTVITHQDVFRNLNLSHLPTSPSPAESIHRIEGLSOKFIY 404  
 DB 335 NIFVTNGQIPSWNLNLPRTVITHQDVFRNLNLSHLPTSPSPAESIHRIEGLSOKFIY 394  
 QY 405 LNDVFMFGKDVWDDFYSHSGQKVLITWVPNCAECGPGSWIKDGYCDKACNNSACDWD 464  
 DB 395 LNDVFMFGKDVWDDFYSHSGQKVLITWVPNCAECGPGSWIKDGYCDKACNNSACDWD 454  
 QY 465 GGDGSGNSGGGRYTAGGGTSGIVGHPWFGGINSVSYCNOGCANSWLADRFCDQACN 524  
 DB 455 GGDGSGNSGGGRYTAGGGTSGIVGHPWFGGINSVSYCNOGCANSWLADRFCDQACN 514  
 QY 525 VLSGCFDAGDCQDHFHELYKVIILLPNOYTHIIIPKGECLPYFSPFAEVAKGVGAYSDNP 594  
 DB 515 VLSGCFDAGDCQDHFHELYKVIILLPNOYTHIIIPKGECLPYFSPFAEVAKGVGAYSDNP 574  
 QY 595 IIRHASIANKWKTIIHLIMHSGMNATTIHFNLTFONTNDEEFKMQITVEVDTRGPKLNST 644  
 DB 575 IIRHASIANKWKTIIHLIMHSGMNATTIHFNLTFONTNDEEFKMQITVEVDTRGPKLNST 634

QY 645 AQGYENLVSPITLLPEAEIILFEDIPKEKRPFKRHDVNSTRAAQEEVKIPLVNIISLLP 704  
 DB 635 AQGYENLVSPITLLPEAEIILFEDIPKEKRPFKRHDVNSTRAAQEEVKIPLVNIISLLP 694  
 QY 705 KDAQSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQHAKIKNOAILITDETNDSLVA 764  
 DB 695 KDAQSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQHAKIKNOAILITDETNDSLVA 754  
 QY 765 POEKQVHKSIILPNSLGVSRERLQRLTFPAVSVKVNQHDQGNPPDLDETARFRVETHTQK 824  
 DB 755 POEKQVHKSIILPNSLGVSRERLQRLTFPAVSVKVNQHDQGNPPDLDETARFRVETHTQK 814  
 QY 825 TTGGNVTKEKPSLIVPLESQMTKEKKTGKEKNSRMEENAEHNGVTEVLLGRKLOHY 884  
 DB 815 TTGGNVTKEKPSLIVPLESQMTKEKKTGKEKNSRMEENAEHNGVTEVLLGRKLOHY 874  
 QY 885 TDSYLGFLPWEKKYFQDLDEEESLKTQLAYFTDSKNTGR 925  
 DB 875 TDSYLGFLPWEKKYFQDLDEEESLKTQLAYFTDSKNTGR 915

RESULT 7  
 ABW01487  
 ID ABW01487 standard; protein; 1199 AA.  
 XX AC ABW01487;  
 XX DT 15-JAN-2004 (first entry)  
 XX DE N-acetylglucosamine-1 (GlcNAc)-phosphotransferase.  
 XX KW Mannose glycoprotein; gene therapy; carbohydrate deficient cell;  
 KW lysosomal storage disease; N-acetylglucosamine-1-phosphotransferase;  
 KW gastrointestinal; enzyme; lectin resistant cell; deoxymannojirimycin;  
 XX kifunensine; glycosylation inhibition.  
 OS Unidentified.  
 XX US2003124652-A1.  
 XX PN 03-JUL-2003.  
 XX PD 21-DEC-2001; 2001US-00023889.  
 XX PF 21-DEC-2001; 2001US-00023889.  
 XX PR (NOVA-) NOVAZYME PHARM INC.  
 XX PA Canfield WM;  
 XX PI WPI; 2003-810984/76.  
 XX DR N-PSDB; ADD62490.  
 XX PT Producing a high mannose glycoprotein for treating lysosomal storage  
 PT disease, comprises culturing the lectin resistant mammalian cell in the  
 PT presence of deoxymannojirimycin and kifunensine.  
 XX PS Claim 8; Page 10-13; 46pp; English.  
 XX CC The invention relates to a method for producing a high mannose  
 CC glycoprotein. The method comprises: introducing and expressing a  
 CC polynucleotide encoding a glycoprotein into a mammalian cell; culturing  
 CC the cell in the presence of a lectin to obtain a lectin resistant cell;  
 CC isolating the cell; culturing the cell in the presence of  
 CC deoxymannojirimycin and kifunensine to inhibit glycosylation of the  
 CC glycoprotein; and collecting the glycoprotein. The invention is useful in  
 CC gene therapy. The method is useful for producing a high mannose  
 CC glycoprotein in a complex carbohydrate deficient cell for treating  
 CC lysosomal storage disease. The present sequence is N-acetylglucosamine-1  
 CC (GlcNAc)-phosphotransferase  
 XX Sequence 1199 AA;  
 SQ



Query Match 94.4%; Score 4630; DB 7; Length 1199;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 45 SRDQYHVLFDSDYRDNIAGSKFQNRCLPMPIDVVTWNGTDLLELLKELQOVRQMEERQ 104  
DB 35 SRDQYHVLFDSDYRDNIAGSKFQNRCLPMPIDVVTWNGTDLLELLKELQOVRQMEERQ 94

QY 105 KAMREILGNKNTTEPTKSKQLECLLTHCIVKPMVLVDPALPANITLKDVPSLYPSFHS 164  
DB 95 KAMREILGNKNTTEPTKSKQLECLLTHCIVKPMVLVDPALPANITLKDVPSLYPSFHS 154

QY 165 SDIFNVAKPNSTNVSVVFDSTKDVEDAHSGLLKGNRSQTVWRGYLTDTKEVPLVLM 224  
DB 155 SDIFNVAKPNSTNVSVVFDSTKDVEDAHSGLLKGNRSQTVWRGYLTDTKEVPLVLM 214

QY 225 QDLAFISGPPPTFKETNQLTKLPENLSKVLLQLYSEASVALLKLNPKDFQELNKOT 284  
DB 215 QDLAFISGPPPTFKETNQLTKLPENLSKVLLQLYSEASVALLKLNPKDFQELNKOT 274

QY 285 KKNMTIDGKELTISPAYLLWDLISQSKODEDISASRFEDNEELRYSLRSIERHAPWR 344  
DB 275 KKNMTIDGKELTISPAYLLWDLISQSKODEDISASRFEDNEELRYSLRSIERHAPWR 334

QY 345 NIFIVTNGQIPSWNLNDNPRVTIVTHQDVFRNLSHLPTFSSPAIESHVRHIEGLSQKFIY 404  
DB 335 NIFIVTNGQIPSWNLNDNPRVTIVTHQDVFRNLSHLPTFSSPAIESHVRHIEGLSQKFIY 394

QY 405 LNDDVMFGKDVWPDFFYSHSGQKQYLLTWVPNCAGCGPGSWIKDGYCDKACNNSACDWD 464  
DB 395 LNDDVMFGKDVWPDFFYSHSGQKQYLLTWVPNCAGCGPGSWIKDGYCDKACNNSACDWD 454

QY 465 GGDCSGNSGGSYIAGGGGTGSGIGVGPWQFGGGINSVSYCNQGCANSWLADKFCDOACN 524  
DB 455 GGDCSGNSGGSYIAGGGGTGSGIGVGPWQFGGGINSVSYCNQGCANSWLADKFCDOACN 514

QY 525 VLSGCFDAGDCQDHFHLYKVILLPNQTHVIIIPKGECLPYFSPAIAEVAKRGEVAGSDNP 584  
DB 515 VLSGCFDAGDCQDHFHLYKVILLPNQTHVIIIPKGECLPYFSPAIAEVAKRGEVAGSDNP 574

QY 585 IIRHASIANKWKTHILMHSGNATTHFNLTFTQNTDDEEPMQITVEVDTRGPKLNST 644  
DB 575 IIRHASIANKWKTHILMHSGNATTHFNLTFTQNTDDEEPMQITVEVDTRGPKLNST 634

QY 645 AQGYENLVSPITILLPEAEILFEDIPKRPFPKFRHDVNSTRAAQBEVKIPLVNIISLP 704  
DB 635 AQGYENLVSPITILLPEAEILFEDIPKRPFPKFRHDVNSTRAAQBEVKIPLVNIISLP 694

QY 705 KDAQLSNTLQLEHGDITLKGYNLKSKSALLRFLMNSQAKIKNOAIIITDETNDLSVA 764  
DB 695 KDAQLSNTLQLEHGDITLKGYNLKSKSALLRFLMNSQAKIKNOAIIITDETNDLSVA 754

QY 765 PQEKQVHKSILPNSLGVSERLQRLTFPAVSVKVGNGHQGNPPLDLETTARFRVETHQK 824  
DB 755 PQEKQVHKSILPNSLGVSERLQRLTFPAVSVKVGNGHQGNPPLDLETTARFRVETHQK 814

QY 825 TIGGNVTKRPPSLIPLVLESQMTKEKKITGKEKNSRMEENAEHIGVTEVLLGRKLQHY 884  
DB 815 TIGGNVTKRPPSLIPLVLESQMTKEKKITGKEKNSRMEENAEHIGVTEVLLGRKLQHY 874

QY 885 TDSYLGFLPWEKKYKFODLDEESLKTOLAYFTDSKNTGR 925  
DB 875 TDSYLGFLPWEKKYKFODLDEESLKTOLAYFTDSKNTGR 915

RESULT 8

ABW01536

ID ABW01536 standard; protein; 1199 AA.

XX AC

XX ABW01536;

XX DT

15-JAN-2004 (first entry)

XX N-acetylglucosamine-1 (GlcNAc)-phosphotransferase protein.  
DE Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;  
KW N-acetylglucosamine-1-phosphotransferase; gene therapy; enzyme.  
XX Unidentified.  
OS  
XX US2003124653-A1.  
PN 03-JUL-2003.  
XX  
PF 21-DEC-2001; 2001US-00023890.  
XX  
PR 21-DEC-2001; 2001US-00023890.  
XX  
PA (NOVA-) NOVAZYME PHARM INC.  
XX Canfield WM;  
PI  
XX WPI; 2003-810985/76.  
DR N-PSDB; AAD62649.  
XX  
XX Producing a glycoprotein with reduced complex carbohydrates by culturing  
PT the lectin resistant mammalian cell expressing the glycoprotein for  
PT treating lysosomal storage disease.  
XX  
PS Claim 8; Page 10-13; 46pp; English.  
XX  
XX The present invention provides a method of producing a glycoprotein  
CC having reduced complex carbohydrates by culturing a glycoprotein  
CC mammalian cell expressing the glycoprotein. The method is useful for  
CC producing a glycoprotein with reduced complex carbohydrates for treating  
CC lysosomal storage disease. The present invention is also useful in gene  
CC therapy. The present sequence is N-acetylglucosamine (GlcNAc)-  
CC phosphotransferase protein  
XX  
SQ Sequence 1199 AA;

Query Match 94.4%; Score 4630; DB 7; Length 1199;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 45 SRDQYHVLFDSDYRDNIAGSKFQNRCLPMPIDVVTWNGTDLLELLKELQOVRQMEERQ 104  
DB 35 SRDQYHVLFDSDYRDNIAGSKFQNRCLPMPIDVVTWNGTDLLELLKELQOVRQMEERQ 94

QY 105 KAMREILGNKNTTEPTKSKQLECLLTHCIVKPMVLVDPALPANITLKDVPSLYPSFHS 164  
DB 95 KAMREILGNKNTTEPTKSKQLECLLTHCIVKPMVLVDPALPANITLKDVPSLYPSFHS 154

QY 165 SDIFNVAKPNSTNVSVVFDSTKDVEDAHSGLLKGNRSQTVWRGYLTDTKEVPLVLM 224  
DB 155 SDIFNVAKPNSTNVSVVFDSTKDVEDAHSGLLKGNRSQTVWRGYLTDTKEVPLVLM 214

QY 225 QDLAFISGPPPTFKETNQLTKLPENLSKVLLQLYSEASVALLKLNPKDFQELNKOT 284  
DB 215 QDLAFISGPPPTFKETNQLTKLPENLSKVLLQLYSEASVALLKLNPKDFQELNKOT 274

QY 285 KKNMTIDGKELTISPAYLLWDLISQSKODEDISASRFEDNEELRYSLRSIERHAPWR 344  
DB 275 KKNMTIDGKELTISPAYLLWDLISQSKODEDISASRFEDNEELRYSLRSIERHAPWR 334

QY 345 NIFIVTNGQIPSWNLNDNPRVTIVTHQDVFRNLSHLPTFSSPAIESHVRHIEGLSQKFIY 404  
DB 335 NIFIVTNGQIPSWNLNDNPRVTIVTHQDVFRNLSHLPTFSSPAIESHVRHIEGLSQKFIY 394

QY 405 LNDDVMFGKDVWPDFFYSHSGQKQYLLTWVPNCAGCGPGSWIKDGYCDKACNNSACDWD 464  
DB 395 LNDDVMFGKDVWPDFFYSHSGQKQYLLTWVPNCAGCGPGSWIKDGYCDKACNNSACDWD 454

QY 465 GGDCSGNSGGSYIAGGGGTGSGIGVGPWQFGGGINSVSYCNQGCANSWLADKFCDOACN 524  
DB 455 GGDCSGNSGGSYIAGGGGTGSGIGVGPWQFGGGINSVSYCNQGCANSWLADKFCDOACN 514

```

Db 455 GDCSGNSGSRRIAGGGTSGIGVGQPMQFOGGINSVSYCNQGCANSWLADKECDQACN 514
Qy 525 VLSGCGDAGDCGDHPEHYKVIILLPNOHYIIPKGECLPYFSFAEAKRGVEGAYSDNP 584
Db 515 VLSGCGDAGDCGDHPEHYKVIILLPNOHYIIPKGECLPYFSFAEAKRGVEGAYSDNP 574
Qy 585 IIRHAIANKWTIHLIMSGMNAATIHFNLTQNTNDEEFKMQITVEVDTRGPKLNST 644
Db 575 IIRHAIANKWTIHLIMSGMNAATIHFNLTQNTNDEEFKMQITVEVDTRGPKLNST 634
Qy 645 AQGYENLVSPITLLPEABILFEDIPKEKRPKFKRHDVNSTRAQAEVKIPLVNISLLP 704
Db 635 AQGYENLVSPITLLPEABILFEDIPKEKRPKFKRHDVNSTRAQAEVKIPLVNISLLP 694
Qy 705 KDAQLSINTLDLQLEHGDITLKGYNLKSALLRSFLMNSQAKIKNQAIITDETNDLSVA 764
Db 695 KDAQLSINTLDLQLEHGDITLKGYNLKSALLRSFLMNSQAKIKNQAIITDETNDLSVA 754
Qy 765 PQEKQVHKSIPLNSLGVSRRLQRLTFPAVSVKNGHDOQNPPDLDTTARFRVETHQK 824
Db 755 PQEKQVHKSIPLNSLGVSRRLQRLTFPAVSVKNGHDOQNPPDLDTTARFRVETHQK 814
Qy 825 TIGGNTKPKPSLIVPLLESQMTKEKKITGKEKENSMEENAEHIGVTEVLLGRKLOHY 884
Db 815 TIGGNTKPKPSLIVPLLESQMTKEKKITGKEKENSMEENAEHIGVTEVLLGRKLOHY 874
Qy 885 TDSYLGFLPWEKKYFQDLDEESLKTQLAYFTDSKNTGR 925
Db 875 TDSYLGFLPWEKKYFQDLDEESLKTQLAYFTDSKNTGR 915

RESULT 9
AAE25294
ID AAE25294 standard; protein; 1196 AA.
AC AAE25294;
XX
DT 30-OCT-2002 (first entry)
XX
DE Human nucleic acid-associated protein (NAAP-13).
XX
KW Human; nucleic acid-associated protein; NAAP-13; neurological disorder;
KW arteriosclerosis; cancer; cell proliferative disorder; atherosclerosis;
KW lymphoma; epilepsy; Alzheimer's disease; developmental; anticonvulsant;
KW autoimmune disorder; AIDS; allergy; anaemia; stroke; malaria; leishmania;
KW gene therapy; nootropic; neuroprotective; cerebroprotective; virucide;
KW immunosuppressive; protozoicide; antimicrobial.
XX
OS Homo sapiens.
XX
FH Key
FH Peptide
FH /label= Signal_peptide
FT Domain
FT /note= "Cytosolic domain"
FT Domain
FT /note= "Transmembrane domain"
FT Protein
FT /note= "Mature human NAAP-13"
FT Domain
FT /note= "Non-cytosolic domain"
FT Domain
FT /note= "Transmembrane domain"
FT Domain
FT /note= "Cytosolic domain"
XX
PN WO200250279-A2.
XX
PD 27-JUN-2002.
XX
PF 19-DEC-2001; 2001WO-US050256.
XX
PR 21-DEC-2000; 2000US-0257714P.

```

```

PR 05-JAN-2001; 2001US-0260081P.
PR 16-JAN-2001; 2001US-0262302P.
PR 23-JAN-2001; 2001US-0263823P.
PR 02-FEB-2001; 2001US-0266088P.
PR 29-OCT-2001; 2001US-0348442P.
XX (INCY-) INCYTE GENOMICS INC.
PA Baughin MR, Lu Y, Arvizu C, Ramkumar J, Yao MG, Policky JL;
XX Wallia NK, Tribouley KM, Yue H, Batra S, Ding L, Lal PG;
PI Borowsky ML, Lu DM, Gandhi AR, Griffin JA, Xu Y, Azimzai Y;
PI Gierzen KJ, Tang YT, Warren BA, Mason PM, Burford N, Hafalia AJA;
PI Lee EA, Yang J, Gorvad AS, Emerling BM, Marquis JP, Lee SY;
PI Swarnakar A, Reddy R;
XX WPI; 2002-519887/55.
DR N-PSDB; AAD41203.
XX
PT Nucleic acid associated proteins and nucleic acids for diagnosing,
PT treating and preventing cell proliferative (e.g. cancers), neurological
PT (e.g. epilepsy or stroke) or autoimmune disorders (e.g. AIDS).
XX
PS Claim 88; Page 169-172; 193pp; English.
XX
CC The invention relates to nucleic acid-associated proteins (NAAP) and
CC nucleic acids. The nucleic acid and amino acid sequences are useful for
CC diagnosing, treating and preventing cell proliferative e.g.
CC arteriosclerosis, atherosclerosis, lymphoma or cancers), neurological
CC (e.g. epilepsy, Alzheimer's disease or stroke), developmental, and
CC autoimmune disorders (e.g. AIDS, allergies, or anaemia) or infections
CC (e.g. malaria, or leishmania), as well as in assessing the effects of
CC exogenous compound on the expression of nucleic acid and amino acid
CC sequences of nucleic acid-associated proteins. The invention is useful in
CC gene therapy. The present sequence is human NAAP-13
XX
SQ Sequence 1196 AA;
XX
Query Match 91.7%; Score 4500; DB 5; Length 1196;
Best Local Similarity 93.4%; Pred. No. 0;
Matches 867; Conservative 1; Mismatches 0; Indels 60; Gaps 1;
Qy 1 MLFKLLQRTYTCLSHRYGLVYVFLGVVTVIVSAFQGEVLEWSDQXHVLPDSYRDNI 60
Db 1 MLFKLLQRTYTCLSHRYGLVYVFLGVVTVIVSAFQGEVLEWSDQXHVLPDSYRDNI 60
Qy 61 AGKSFQNRCLPMPIDIVVYTWNGTDLLELKELOVREQMEEECKAMRILGKNTTEPTK 120
Db 61 AGKSFQNRCLPMPIDIVVYTWNGTDLLELKELOVREQMEEECKAMRILGKNTTEPTK 120
Qy 121 KSEKQLECLLTHCIKVPMLVLDLPALPANITLKDVPSPSYFSFHSASDIFNVAKPKNSTNV 180
Db 121 KSEKQLECLLTHCIKVPMLVLDLPALPANITLKDVPSPSYFSFHSASDIFNVAKPKNSTNV 180
Qy 181 SVVVPDSTKQVEDAHSGLLKGNRQTVWRGYLTDDKEVPCVLVMDLAFSLGPPPTFKET 240
Db 181 SVVVPDSTKQVEDAHSGLLKGNRQTVWRGYLTDDKEVPCVLVMDLAFSLGPPPTFKET 240
Qy 241 NQLTKLPENLSSKVKLLQLYSEASVALLKLNPKDPQELNKKOTKKNMTIDGKELTISA 300
Db 241 NQLTKLPENLSSKVKLLQLYSEASVALLKLNPKDPQELNKKOTKKNMTIDGKELTISA 300
Qy 301 YLLWDLISAISQSKODEDISASRPEDNEELYSURSTERHAPWVRNIFVTNGOIPSWLNL 360
Db 301 YLLWDLISAISQSKODEDISASRPEDNEELYSURSTERHAPWVRNIFVTNGOIPSWLNL 360
Qy 361 DNPRVTIVTHQDVFNRLSHLPTFSSPAIESHIHRIEGLSQKFTYLLNDVWFGKDVWPDF 420
Db 361 DNPRVTIVTHQDVFNRLSHLPTFSSPAIESHIHRIEGLSQKFTYLLNDVWFGKDVWPDF 420
Qy 421 YSHSKGQKVTLPVFNCAEGCTGWSIKDGYCDKACNACDWDGDCSGNSGGRYING 480
Db 421 YSHSKGQKVTLPVFNCAEGCTGWSIKDGYCDKACNACDWDGDCSGNSGGRYING 480

```

QY 481 GGGTSGIGVGHWPQFGGGINSVYCNQGCANSLWADKFCDOACNVLSGCFDAGCGQDHF 540  
 Db 435 -----FGGGINSVYCNQGCANSLWADKFCDOACNVLSGCFDAGCGQDHF 480  
 QY 541 HELYKVILLPNQTHYIIIPKGECLPYFSPAFAVAKRGEVAGSDNPIIRHASIANKWKTIHL 600  
 Db 481 HELYKVILLPNQTHYIIIPKGECLPYFSPAFAVAKRGEVAGSDNPIIRHASIANKWKTIHL 540  
 QY 601 IMHSGMNATTHFNLTFTQNTNDEEFKQITVEVDTRGPKLNSTAKGYENLVSPITLLP 660  
 Db 541 IMHSGMNATTHFNLTFTQNTNDEEFKQITVEVDTRGPKLNSTAKGYENLVSPITLLP 600  
 QY 661 EAEILFEDIPKEKRPFKRHDVNSTRAAQEVEKIPLVNISLLPKDAQLSNTLDLQLEH 720  
 Db 601 EAEILFEDIPKEKRPFKRHDVNSTRAAQEVEKIPLVNISLLPKDAQLSNTLDLQLEH 660  
 QY 721 GDITLKGYNLSKALLRFLMNSQHAKIKNOAIIITDETNDLSLVAPOEKQVHKSILPNSLG 780  
 Db 661 GDITLKGYNLSKALLRFLMNSQHAKIKNOAIIITDETNDLSLVAPOEKQVHKSILPNSLG 720  
 QY 781 VSERLQRLTFPAVSVKVNHDGQGNPPDLLETTARFRVETHQKTIGGNVTKKEPPSLIV 840  
 Db 721 VSERLQRLTFPAVSVKVNHDGQGNPPDLLETTARFRVETHQKTIGGNVTKKEPPSLIV 780  
 QY 841 PLESQMTKEKKITGKEKENSMEENAENHIGVTEVLLGRKLOHYTDSYLGFLPWEKKKYF 900  
 Db 781 PLESQMTKEKKITGKEKENSMEENAENHIGVTEVLLGRKLOHYTDSYLGFLPWEKKKYF 840  
 QY 901 QDLDEESLTKQLAYFTDSKNTGRQLK 928  
 Db 841 QDLDEESLTKQLAYFTDSKNTGRQLK 868

RESULT 10  
 ADA54824  
 ID ADA54824 standard; protein; 846 AA.  
 XX AC ADA54824;  
 XX DT 20-NOV-2003 (first entry)  
 XX DE Human protein, SEQ ID 2392.  
 XX KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;  
 KW Gene therapy; human; secretory protein; membrane proteins; cancer;  
 KW inflammatory disease; osteoporosis; neurological disease.  
 XX OS Homo sapiens.  
 XX PN EP1293569-A2.  
 XX PD 19-MAR-2003.  
 XX PF 21-MAR-2002; 2002EP-00006586.  
 XX PR 14-SEP-2001; 2001JP-00328381.  
 XX ER 24-JAN-2002; 2002US-0350435P.  
 XX PA (HELI-) HELIX RES INST.  
 XX PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
 XX PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,  
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,  
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;  
 XX WPI; 2003-395539/38.  
 XX DR N-PSDB; ADA53185.  
 XX PT New polynucleotides encoding full-length polypeptides, e.g. secretory  
 PT and/or membrane proteins, useful for developing medicines for diseases in  
 PT which the gene is involved, or as target molecules for gene therapy.  
 XX PS Claim 14; SEQ ID NO 2392; 205pp; English.

XX The present invention relates to novel human secretory or membrane  
 CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
 CC ADA54071). The coding sequences are useful in the gene therapy of  
 CC diseases caused by abnormalities of the proteins, e.g. cancer,  
 CC inflammatory diseases, osteoporosis or neurological disease.  
 XX

XX Sequence 846 AA;

Query Match 91.1%; Score 4468; DB 6; Length 846;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 844; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLEKLLQRTYTCLSHRYGLYVCFGLGVVTVIYSAQFGVEVLEWSDQVHVLPDSYRDM 60

Db 1 MLEKLLQRTYTCLSHRYGLYVCFGLGVVTVIYSAQFGVEVLEWSDQVHVLPDSYRDM 60

QY 61 AGSFQNRCLPMPIDVVTWNGTDELLKELQVREOMESEKAMREILGNVTEPTK 120

Db 61 AGSFQNRCLPMPIDVVTWNGTDELLKELQVREOMESEKAMREILGNVTEPTK 120

QY 121 KSEKQLECLLTHCIKVPMLVLPALPANITLKDVPSPYPSFHSASDIFNVAKPQSTNV 180

Db 121 KSEKQLECLLTHCIKVPMLVLPALPANITLKDVPSPYPSFHSASDIFNVAKPQSTNV 180

QY 181 SVVVPDSTKDVEDAHSGLLKGNRSQTVWRGYLTTOKEVPLVLMODLAFSLGPPPTFKET 240

Db 181 SVVVPDSTKDVEDAHSGLLKGNRSQTVWRGYLTTOKEVPLVLMODLAFSLGPPPTFKET 240

QY 241 NQKTKLPENLSKVKLQLYSEASVALLKLNPNKDFQELNKQTKKNMTIDGKELTISPA 300

Db 241 NQKTKLPENLSKVKLQLYSEASVALLKLNPNKDFQELNKQTKKNMTIDGKELTISPA 300

QY 301 YLLWDLAISQSKQDEDISASRPEDNEELRYSLRISIERHAPVWRNIFIVTNGQIPSWNL 360

Db 301 YLLWDLAISQSKQDEDISASRPEDNEELRYSLRISIERHAPVWRNIFIVTNGQIPSWNL 360

QY 361 DNPRVTIVTHQDVFRNLSHLPTFFSPAIESH;HRIEGLSOKETIYLNDDVWFGKDVWPDF 420

Db 361 DNPRVTIVTHQDVFRNLSHLPTFFSPAIESH;HRIEGLSOKETIYLNDDVWFGKDVWPDF 420

QY 421 YSHSGKQKYLTPVPNCABGCPGSKIWDGYCDKACNNSACDWDGDCSGNSGSGRYIAG 480

Db 421 YSHSGKQKYLTPVPNCABGCPGSKIWDGYCDKACNNSACDWDGDCSGNSGSGRYIAG 480

QY 481 GGGTSGIGVGHWPQFGGGINSVYCNQGCANSLWADKFCDOACNVLSGCFDAGCGQDHF 540

Db 481 GGGTSGIGVGHWPQFGGGINSVYCNQGCANSLWADKFCDOACNVLSGCFDAGCGQDHF 540

QY 541 HELYKVILLPNQTHYIIIPKGECLPYFSPAFAVAKRGEVAGSDNPIIRHASIANKWKTIHL 600

Db 541 HELYKVILLPNQTHYIIIPKGECLPYFSPAFAVAKRGEVAGSDNPIIRHASIANKWKTIHL 600

QY 601 IMHSGMNATTHFNLTFTQNTNDEEFKQITVEVDTRGPKLNSTAKGYENLVSPITLLP 660

Db 601 IMHSGMNATTHFNLTFTQNTNDEEFKQITVEVDTRGPKLNSTAKGYENLVSPITLLP 660

QY 661 EAEILFEDIPKEKRPFKRHDVNSTRAAQEVEKIPLVNISLLPKDAQLSNTLDLQLEH 720

Db 661 EAEILFEDIPKEKRPFKRHDVNSTRAAQEVEKIPLVNISLLPKDAQLSNTLDLQLEH 720

QY 721 GDITLKGYNLSKALLRFLMNSQHAKIKNOAIIITDETNDLSLVAPOEKQVHKSILPNSLG 780

Db 721 GDITLKGYNLSKALLRFLMNSQHAKIKNOAIIITDETNDLSLVAPOEKQVHKSILPNSLG 780

QY 781 VSERLQRLTFPAVSVKVNHDGQGNPPDLLETTARFRVETHQKTIGGNVTKKEPPSLIV 840

Db 781 VSERLQRLTFPAVSVKVNHDGQGNPPDLLETTARFRVETHQKTIGGNVTKKEPPSLIV 840

QY 841 PLESQMTKEKKITGKEKENSMEENAENHIGVTEVLLGRKLOHYTDSYLGFLPWEKKKYF 900

Db 841 PLESQMTKEKKITGKEKENSMEENAENHIGVTEVLLGRKLOHYTDSYLGFLPWEKKKYF 900

RESULT 11  
ADG39856  
ID ADG39856 standard; protein; 847 AA.  
XX AC ADG39856;  
XX DT 26-FEB-2004 (first entry)  
XX DE Protein similar to human NOV12 #2.  
XX KW Human; NOVX; cardiomyopathy; atherosclerosis; hypertension;  
KW congenital heart defect; pulmonary stenosis; scleroderma; obesity;  
KW metabolic disturbance; obesity; transplantation; adrenoleukodystrophy;  
KW congenital adrenal hyperplasia; prostate cancer; diabetes;  
KW metabolic disorder; neoplasm; adenocarcinoma; fertility; haemophilia;  
KW graft versus host disease; AIDS; bronchial asthma; Crohn's disease;  
KW multiple sclerosis; infectious disease; anorexia;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW immune disorder; haematopoietic disorder; dyslipidaemia;  
XX KW wasting disorder.  
XX OS Homo sapiens.  
XX PN US2003203843-A1.  
XX PD 30-OCT-2003.  
XX PF 11-APR-2002; 2002US-00120801.  
XX PR 20-APR-2001; 2001US-0285609P.  
XX PR 23-APR-2001; 2001US-0285748P.  
XX PR 24-APR-2001; 2001US-0286068P.  
XX PR 25-APR-2001; 2001US-0286292P.  
XX PR 03-MAY-2001; 2001US-0288334P.  
XX PR 16-MAY-2001; 2001US-0291241P.  
XX PR 14-SEP-2001; 2001US-0322284P.  
XX PA (PENA/) PENA C E A.  
XX PA (GUOX/) GUO X.  
XX PA (SHIM/) SHIMKETS R A.  
XX PA (PADI/) PADIGARU M.  
XX PA (KEKU/) KEKUDA R.  
XX PA (SEYI/) SEYTEK K A.  
XX PA (MEHR/) MEHRABAN F.  
XX PA (TOPP/) TOPPER J N.  
XX PA (MALY/) MALYANKAR U M.  
XX PA (WASS/) WASSERMAN S M.  
XX PA (EDIN/) EDINGER S R.  
XX PA (SMIT/) SMITHSON G.  
XX PA (GUNT/) GUNTHER E.  
XX PA (KOMU/) KOMUVES L.  
XX PI PENA CEA, Guo X, Shinkets RA, Padigaru M, Kekuda R, Spytek KA;  
PI Mehraban F, Topper JN, Malvankar UM, Wasserman SM, Edinger SR;  
PI Smithson G, Gunther E, Komuves L;  
XX WPI; 2003-900671/82.  
XX DR New NOVX polypeptides and nucleic acids, useful for diagnosing or  
XX PT treating e.g. cardiomyopathy, atherosclerosis, hypertension, scleroderma,  
XX PT obesity, prostate cancer, AIDS, bronchial asthma, Crohn's disease, or  
XX PT multiple sclerosis.  
XX PS Disclosure; SEQ ID NO 88; 215pp; English.  
XX CC The invention relates to a new isolated polypeptide comprising an amino  
XX CC acid sequence selected from 17 fully defined human NOVX sequences (even  
XX CC SEQ ID NOS between ADG39770 and ADG39802), a mature form of the NOVX  
XX CC amino acid or a variant of NOVX, where one or more amino acid residue in  
XX CC the variant differs in no more than 15% of the amino acid residues of  
XX CC NOVX. Also included are an isolated nucleic acid (NA) molecule  
XX CC (comprising a nucleic acid sequence encoding a NOVX polypeptide above

(odd SEQ ID NOS between ADG39769 and ADG39801), a nucleic acid fragment encoding at least a portion of a NOVX polypeptide and a complement of NOVX NA), a vector comprising NOVX NA, a cell comprising the vector, an antibody that immunospecifically binds to NOVX, a method for determining the presence or amount of NOVX or NOVX NA in a sample, a method of identifying an agent that binds to NOVX, a method for identifying an agent that modulates the expression or activity of NOVX, a method for modulating the activity of NOVX, a method of treating or preventing a NOVX-associated disorder, a method for screening for a modulator of activity or of latency or predisposition to a NOVX-associated disorder, a method for determining the presence of or predisposition to a disease associated with altered levels of NOVX or NOVX NA in a first mammalian subject and a method of treating a pathological state in a mammal by administering NOVX or an antibody that binds to NOVX. The NOVX polypeptide, nucleic acid or antibody is useful in the manufacture of a medicament for treating a syndrome associated with a human disease or a NOVX-associated disorder. The NOVX polypeptides and nucleic acids encoding them are useful for diagnosing or treating pathologies, diseases or conditions associated with NOVX sequences, including cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, pulmonary stenosis, scleroderma, obesity, metabolic disturbances associated with obesity, transplantation, adrenoleukodystrophy, congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic disorders, neoplasm, adenocarcinoma, fertility, haemophilia, graft versus host disease, AIDS, disease, anorexia, neurodegenerative disorders (e.g. Alzheimer's disease, or Parkinson's disease), immune disorders, haematopoietic disorders, dyslipidaemias, and wasting disorders associated with chronic diseases. The polypeptides can be used as immunogens to produce antibodies and as vaccines. The sequences may further be used in chromosome mapping, identifying individual from minute biological samples (tissue typing), and in forensic identification of a biological sample. The present sequence is a protein showing sequence similarity to a NOVX protein.

Sequence 847 AA;  
Query Match 90.9%; Score 4461; DB 7; Length 847;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 842; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MLFKLLQRTYTCLSHRYGLVYVCFVGVVTVIVSAFQGEVLEWRSRQYHVLFDSDYRDI 60  
DB 1 MLFKLLQRTYTCLSHRYGLVYVCFVGVVTVIVSAFQGEVLEWRSRQYHVLFDSDYRDI 60  
QY 61 AKSFQNLCLPMPIDVYVTVWNGTDLLELKLQVREQMEEEKAMREILGKNTTPTK 120  
DB 61 AKSFQNLCLPMPIDVYVTVWNGTDLLELKLQVREQMEEEKAMREILGKNTTPTK 120  
QY 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKQVPSLYPSFHSASDIFNVAKPKNPTNV 180  
DB 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKQVPSLYPSFHSASDIFNVAKPKNPTNV 180  
QY 181 SVVFPDSTKVEDAHSGLLKGNRQTVWRGYLTDDKEVPGVLVMDLAFILSGFPPTKET 240  
DB 181 SVVFPDSTKVEDAHSGLLKGNRQTVWRGYLTDDKEVPGVLVMDLAFILSGFPPTKET 240  
QY 241 NQKTKLPENLSKVKLLQLYSEASVALLKLNPKDFQELNKTQKNTLDGKELTISPA 300  
DB 241 NQKTKLPENLSKVKLLQLYSEASVALLKLNPKDFQELNKTQKNTLDGKELTISPA 300  
QY 301 YLLWDLSSAISQSKODETISASRFEDNEELRYSLRSIERHAPWVRNIFIVNGQIPSWLNL 360  
DB 301 YLLWDLSSAISQSKODETISASRFEDNEELRYSLRSIERHAPWVRNIFIVNGQIPSWLNL 360  
QY 361 DNPRTVITVHODVFNRLSHLPTFSSPAIESHIEGLSQKFIYLNDDVWFKDWWDDF 420  
DB 361 DNPRTVITVHODVFNRLSHLPTFSSPAIESHIEGLSQKFIYLNDDVWFKDWWDDF 420  
QY 421 YSHSKGQKYLITWVPNCAEGCPGSIKDGCDKACNNACDWDGDCSGNSGGSRYIAG 480  
DB 421 YSHSKGQKYLITWVPNCAEGCPGSIKDGCDKACNNACDWDGDCSGNSGGSRYIAG 480  
QY 481 GGGTSGIGVHPWQFGGGINSVSYCNQCANSLADKFCQACNVLSGDFDAGDGGQDHF 540

Db 481 GGGTSGVGPWQGGGINSVYCNQGCANSLADKFCDCQACNLSGFGDAGCGQDHF 540  
QY 541 HELYKVILLPNQTHYIIIPKGECLPYFSAFAVAKRGVEGAYSDNPIIRHASIANKWKTHL 600  
Db 541 HELYKVILLPNQTHYIIIPKGECLPYFSAFAVAKRGVEGAYSDNPIIRHASIANKWKTHL 600  
QY 601 IMHSGWNTTTHFNLTONTNDEEFKMOITVEVDTRFGPKLNSTAKGYENLVSPITLLP 660  
Db 601 IMHSGWNTTTHFNLTONTNDEEFKMOITVEVDTRFGPKLNSTAKGYENLVSPITLLP 660  
QY 661 EAEILFEDIPKEKPFKRDHVNSTRQAQBEVKIPLVNIISLLPKDAQLSNTLDLOLEH 720  
Db 661 EAEILFEDIPKEKPFKRDHVNSTRQAQBEVKIPLVNIISLLPKDAQLSNTLDLOLEH 720  
QY 721 GDITLKGYNLSKALLRFLMNSQAKIKNQAITDETNDLSLVAPOEKVHKSLPNSLG 780  
Db 721 GDITLKGYNLSKALLRFLMNSQAKIKNQAITDETNDLSLVAPOEKVHKSLPNSLG 780  
QY 781 VSERLQRLTTPFAVSVKVNGHDQGNPDLDTTARFRVETHTQKTIGNVTKKPPSLIV 840  
Db 781 VSERLQRLTTPFAVSVKVNGHDQGNPDLDTTARFRVETHTQKTIGNVTKKPPSLIV 840  
QY 841 PLESQMT 847  
Db 841 PLESQMT 847

RESULT 12

ABU07381 ID ABU07381 standard; protein; 1459 AA.

XX AC

XX ABU07381;

XX DT 28-JAN-2003 (first entry)

XX DE Human protein NOV12.

XX KW Human; NOVX; cardiomyopathy; atherosclerosis; cancer; hypertension;  
XX KW diabetes; inflammation; autoimmune disorder; allergy; blood disorder;  
XX KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;  
XX KW immunoglobulin A nephropathy; cirrhosis; arthritis; Alzheimer's disease;  
XX KW Parkinson's disease; goitre; infection; stroke; muscular dystrophy;  
XX KW epilepsy; wasting disorder; neurogenesis; cell differentiation;  
XX KW cell proliferation; haematopoiesis; wound healing; angiogenesis;  
XX KW gene therapy; single nucleotide polymorphism; SNP.

XX OS Homo sapiens.

XX EN WO200285922-A2.

XX FD 31-OCT-2002.

XX PF 11-APR-2002; 2002WO-US011634.

XX PR 23-APR-2001; 2001US-0285748P.

XX PR 24-APR-2001; 2001US-0286068P.

XX PR 25-APR-2001; 2001US-0286292P.

XX PR 03-MAY-2001; 2001US-0288334P.

XX PR 16-MAY-2001; 2001US-0291241P.

XX PR 14-SEP-2001; 2001US-0322284P.

XX PA (CURA-) CURAGEN CORP.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX PI Pena CEA, Guo X, Shinkets RA, Padigaru M, Kekuda R, Spytek KA;

XX PI Mehrahan F, Topper JN, Malyankar UM, Wasserman S, Edinger S;

XX PI Smithson G, Gunther E, Komuves L;

XX XX WPI; 2003-058712/05.

XX DR N-PSDB; ABX10235.

XX XX New NOVX polypeptides and nucleic acids, useful for preventing or

PT treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,  
PT atherosclerosis or diabetes, and in chromosome mapping, tissue typing or  
PT pharmacogenomics.

PS Claim 1; Page 94-95; 301pp; English.

XX The invention relates to an isolated polypeptide comprising any one of 17  
XX human NOVX (1-9, 10a, 10b, 11-16) appearing as ABU07369-ABU07385, a  
XX mature form of it, or a variant of them, where one or more residues of  
XX the variant differs in not more than 15 % from the residues of the  
XX sequence of them and their encoding polynucleotides appearing as ABX10223  
XX -ABX10239. Also included are NOVX expression vectors, transfected cells,  
XX antibodies, identifying an agent that binds to or modulates the  
XX expression or activity of NOVX and screening for a modulator of activity  
XX or of latency or predisposition to a NOVX-associated disorder. The NOVX  
XX polypeptides, polynucleotides and antibodies are useful in manufacturing  
XX a medicament for treating or preventing a syndrome associated with NOVX-  
XX associated disorder, such as cardiomyopathy, atherosclerosis, cancer,  
XX hypertension, diabetes, inflammation, autoimmune disorders, allergies,  
XX blood disorders, acquired immunodeficiency syndrome (AIDS), obesity,  
XX asthma, immunoglobulin (Ig)A nephropathy, cirrhosis, arthritis,  
XX Alzheimer's disease, Parkinson's disease, goitre, infections (e.g.  
XX bacterial, viral, parasitic), stroke, muscular dystrophy, epilepsy, and  
XX other wasting disorders associated with chronic diseases. The nucleic  
XX acids and polypeptides may also be used as targets for the identification  
XX of small molecules that modulate or inhibit e.g. neurogenesis, cell  
XX differentiation, cell proliferation, haematopoiesis, wound healing and  
XX angiogenesis, in gene therapy, in generation of antibodies that bind  
XX immunospecifically to NOVX substances for use in therapeutic or  
XX diagnostic methods. The nucleic acids are further used as hybridisation  
XX probes, in chromosome mapping, tissue typing, preventive medicine, and  
XX pharmacogenomics. The polypeptides are also useful as vaccines. The  
XX present sequence represents a NOVX protein

XX SQ Sequence 1459 AA;

Query Match 89.5%; Score 4391; DB 6; Length 1459;

Best Local Similarity 73.0%; Pred. No. 0;

Matches 884; Conservative 13; Mismatches 28; Indels 286; Gaps 13;

QY 1 MLFKLLQRTYTCLSHRYGLYVCFLGVVTVIVSAFQFGE----- 39

Db 1 MLFKLLQRTYTCLSHRYGLYVCFLGVVTVIVSAFQFGEWEARDPAKHPVHTAPTATK 60

QY 40 ----- 39

Db 61 NHPAQNVDSAEVKEGIRGKNGCRAVSLQDWPGRGCANFTFAFCHCKSEVSKRPL 120

QY 40 -----VYLEWSRDQYHVLFDSDYRD 58

Db 121 YILQNCWLTWDGWTWLTALLHGSLLIQGPASEPGCVLLKAKVLEWSRDQYHVLFDSDYRD 180

QY 59 NIAGKSFQNRCLCPMPIDVVYTWVNGTDLLELLKELOVREQMEEEQKAMRETLGKNTTTP 118

Db 181 NIAGKSFQNRCLCPMPIDVVYTWVNGTDLLELLKELOVREQMEEEQKAMRETLGKNTTTP 240

QY 119 TKKSEKQLECLTHCICKVPMVLDPALPANITLKDVPISLPSFHSASDIFNVAKPKNPST 178

Db 241 TKKSEKQLECLTHCICKVPMVLDPALPANITLKDVPISLPSFHSASDIFNVAKPKNPST 300

QY 179 NVSVVVFSTKQ-----VEDAHSLLKGN-----RQTVW----- 208

Db 301 NVSVVVFSTKQDGTLLTQKVTFEWKCEGEVASNANIGWTKDLSFRFLPWPVLEBPR 360

QY 209 -----RGYLTDDKEVPCGLVL-MODLAFSLGFP 234

Db 361 AQLSSALQILTFPRVSDRANTSVEIKLDTPLLRGYA---KPVGPEFGLQPLSAHCLP 417

QY 235 -----PTF---KET-----NQLKT---KL 247

Db 418 TLDRKVNELRDFVKNYKQDPSILHTKETCFREQVESMGESYKSEENIKELATGSKV 477

QY 248 PEN-----LSSKVKLLQLYS-----EASVALLKLNPKDFQELNKQTK 285

478	DB	ENISTDELSSBESLEIDNEAVIPDPTDTSPOEMGDGEASVALLKLNPNKDFQEAJNKOTK	537
286	QY	KNMTTDGKELTISPAVLLWDLISAIQSQODEDISASRPEDNEBELYSLRSIERHAPWVN	345
538	DB	KNMTTDGKELTISPAVLLWDLISAIQSQODEDISASRPEDNEBELYSLRSIERHAPWVN	597
346	QY	IPIVNGQIPSWNLNDNPRVITVTHQDVPFNLNLSLPTSSPAIESHIHRIEGLSKFIYL	405
598	DB	IPIVNGQIPSWNLNDNPRVITVTHQDVPFNLNLSLPTSSPAIESHIHRIEGLSKFIYL	657
406	QY	NDDVWFGKDVWDDPYSHSKGQKLYLTWPVPCACGCGSWTKDGYCDKACNNSACDWDG	465
658	DB	NDDVWFGKDVWDDPYSHSKGQKLYLTWPVPCACGCGSWTKDGYCDKACNNSACDWDG	717
466	QY	GDCSGNSGGSRYIAGGGGTGSGVGHPPWFGGGGINSVSYCNQGCANSWLADFCDOACNV	525
718	DB	GDCSGNSGGSRYIAGGGGTGSGVGHPPWFGGGGINSVSYCNQGCANSWLADFCDOACNV	777
526	QY	LSGCFDAGDCGQ-----DHFPELYKVILLPNQTHYII	557
778	DB	LSGCFDAGDCGQENSCKNRKTEKCPVKKKXIMFLFPFLDHFPELYKVILLPNQTHYII	837
558	QY	PKGECLPYSPFAVAKRGVEGAYSDNPIIRHASIANKWKTIIHLVHSGMNATTIHFNLTF	617
838	DB	PKGECLPYSPFAVAKRGVEGAYSDNPIIRHASIANKWKTIIHLVHSGMNATTIHFNLTF	897
618	QY	QNTNDEEFKQITVEVDTRGPKLNSTAQGYENLVSPITLLPEABILPEDIPEKRPFPK	677
898	DB	QNTNDEEFKQITVEVDTRGPKLNSTAQGYENLVSPITLLPEABILPEDIPEKRPFPK	957
678	QY	FKRHVDNSTRAQEEVKIPLVNIISLLPKDAQSLNLTLDLQLEHGDITLLKGYNLKSALLR	737
958	DB	FKRHVDNSTRAQEEVKIPLVNIISLLPKDAQSLNLTLDLQLEHGDITLLKGYNLKSALLR	1017
738	QY	SFLMNSQHAKIKNQAIITDETNDLSVAPOEKVHKHSILFNSLGVSRRLQRLTFPAVSVKV	797
1018	DB	SFLMNSQHAKIKNQAIITDETNDLSVAPOEKVHKHSILFNSLGVSRRLQRLTFPAVSVKV	1077
798	QY	NGHDQGNPDLLETTARFVETHTKTIGGNVTKKPPSLIVPLESOMTKKKTGKSK	857
1078	DB	NGHDQGNPDLLETTARFVETHTKTIGGNVTKKPPSLIVPLESOMTKKKTGKSK	1137
858	QY	ENSRMEANAENHIGVTEVLLGRKLQHYTDSYLCFLPWEKKYTFQDILLDEESLKTQAYF	917
1138	DB	ENSRMEANAENHIGVTEVLLGRKLQHYTDSYLCFLPWEKKYTFQDILLDEESLKTQAYF	1197
918	QY	TDSKNTGRQLK 928	
1198	DB	TDSKNTGRQLK 1208	
RESULT 13			
ADG39794			
ID	ADG39794 standard; protein; 1459 AA.		
XX	AC		
XX	AC		
XX	AC		
DT	26-FEB-2004 (first entry)		
XX	Human novel protein NOV12.		
DE	XX		
XX	Human; NOVX; cardiomyopathy; atherosclerosis; hypertension;		
KW	congenital heart defect; pulmonary stenosis; scleroderma; obesity;		
KW	metabolic disturbance; obesity; transplantation; adrenoleukodystrophy;		
KW	congenital adrenal hyperplasia; prostate cancer; diabetes;		
KW	metabolic disorder; neoplasm; adenocarcinoma; fertility; haemophilia;		
KW	graft versus host disease; AIDS; bronchial asthma; Crohn's disease;		
KW	multiple sclerosis; infectious disease; anorexia;		
KW	neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;		
KW	immune disorder; haematopoietic disorder; dyslipidaemia;		
XX	wasting disorder.		
XX			



CC obesity, transplantation, adrenoleukodystrophy, congenital adrenal  
CC hyperplasia, prostate cancer, diabetes, metabolic disorders, neoplasm,  
CC adenocarcinoma, fertility, haemophilia, graft versus host disease, AIDS,  
CC bronchial asthma, Crohn's disease, multiple sclerosis, infectious  
CC disease, anorexia, neurodegenerative disorders (e.g. Alzheimer's disease,  
CC or Parkinson's disease), immune disorders, haematopoietic disorders,  
CC dyslipidaemias, and wasting disorders associated with chronic diseases.  
CC The polypeptides can be used as immunogens to produce antibodies and as  
CC vaccines. The sequences may further be used in chromosome mapping,  
CC identifying individual from minute biological samples (tissue typing),  
CC and in forensic identification of a biological sample. The present  
CC sequence represents a NOVX protein.

XX SQ

Query Match

Best Local Similarity 89.5%; Score 4391; DB 7; Length 1459;

Matches 884; Conservative 13; Mismatches 28; Indels 286; Gaps 13;

QY 1 MLFKLQRTYTCLSHRYGLVCFGLGVVTVTSVFQGE----- 39  
DB 1 MLFKLQRTYTCLSHRYGLVCFGLGVVTVTSVFQGEWEARDPAKPIVHRTATYK 60  
QY 40 ----- 39

DB 61 NHPAQNVDSEAEKSGIRRGKNGCRVSLQDWPGRGCAFTAFCHDKCFSEVSQKRFL 120

QY 40 -----VLEWRSRDQVHVLFDSDYRD 58  
DB 121 YILQCHWLTDGWTWALLHGSLLLOGPASEPCVLLKXVLEWRSRDQVHVLFDSDYRD 180

QY 59 NIAGKSFQNLCLPMPIDVTVTVNGTDLLELLKLOQVREMEBEQKAMREILGKNTTEP 118  
DB 181 NIAGKSFQNLCLPMPIDVTVTVNGTDLLELLKLOQVREMEBEQKAMREILGKNTTEP 240

QY 119 TKSEKOLECLLTHCIKVPMLVLDPALPANITLKDVPSLYPSFHSASDIFNVAKPKNPST 178  
DB 241 TKSEKOLECLLTHCIKVPMLVLDPALPANITLKDVPSLYPSFHSASDIFNVAKPKNPST 300

QY 179 NVSVVVFDSKTD-----VEDAHSGLLKGN-----RQTW----- 208  
DB 301 NVSVVVFDSKTDGTLTKQVTFEWEKCEGEVSNANIGKTDLGSPERPLPWPVLEPPR 360

QY 209 -----RGYVTTDKVEPGLV-NQDLAFSGFP 234  
DB 361 AOLSSALQILTRPVSDRANTSVEIKLDTPLLAGYA---KPVGPETGLQPLSFARCLP 417

QY 235 -----PTF---XET-----NQLKT---KL 247  
DB 418 TDLRKVNELRDFVKKYKQDPSILHTKTCFLREQVESMGESYKSEENIKELTKSKKV 477

QY 248 PEN-----LSKVKLLQLYS-----EASVALLKLNPKDFQELNKKQTK 285  
DB 478 EENISTDELSEESDLBIDNEAVIEPDTDSPOEMDGEASVALLKLNPKDFQELNKKQTK 537

QY 286 KNMTIDCKELTISPAVLLWDLAISQSKQDEDISASFEDNEELRYSLRSIERHAPWRN 345  
DB 538 KNMTIDCKELTISPAVLLWDLAISQSKQDEDISASFEDNEELRYSLRSIERHAPWRN 597

QY 346 IFIVTNGQIPSWLNDLPRVTIVTHQVFNRLSHLPTFSSPAIESHIHRIEGLSKFIYL 405  
DB 598 IFIVTNGQIPSWLNDLPRVTIVTHQVFNRLSHLPTFSSPAIESHIHRIEGLSKFIYL 657

QY 406 NDDVMFGKVDKDDFYSHSGKQKYLTPVPCNACGCPGSKIWDGCDKACNNACSDWDG 465  
DB 658 NDDVMFGKVDKDDFYSHSGKQKYLTPVPCNACGCPGSKIWDGCDKACNNACSDWDG 717

QY 466 GDCSGNSGGSGRYAGGGGTSGIGVHPWQFGGINSVYCNQGCANSLWADKFCDOACNV 525  
DB 718 GDCSGNSGGSGRYAGGGGTSGIGVHPWQFGGINSVYCNQGCANSLWADKFCDOACNV 777

QY 526 LSCGFDAGDCGQ-----DHPHELYKVVILLPNQTHYII 557  
DB 526 LSCGFDAGDCGQ-----DHPHELYKVVILLPNQTHYII 557

DB 778 LSCGFDAGDCGQENSDSKNRKTEKCPVKKKIMLFFFLDHPHELYKVVILLPNQTHYII 837  
QY 558 PKGECPLPYTFAEVAKRGVEGAYSDNPIIRHASIANKWKTIHLIMHSGNATTHFNLT 617  
DB 838 PKGECPLPYTFAEVAKRGVEGAYSDNPIIRHASIANKWKTIHLIMHSGNATTHFNLT 857  
QY 618 QNTNDEEFKQIITVEVDTRREGPKLNSTAGKYENLVSPITLLPEABILDIPKEKRF 677  
DB 898 QNTNDEEFKQIITVEVDTRREGPKLNSTAGKYENLVSPITLLPEABILDIPKEKRF 957  
QY 678 FKEDHVNSTRRAQEEVKIPLVNIISLLPKDAQSLNLTDLQLEHGDITLKGYNLSKALL 737  
DB 958 FKEDHVNSTRRAQEEVKIPLVNIISLLPKDAQSLNLTDLQLEHGDITLKGYNLSKALL 1017  
QY 738 SFLMNSCHAKIKNQAIITDETNDLSVAPOEKVHKSILPNSLGVSRLOLQRTFFPAVSV 797  
DB 1018 SFLMNSCHAKIKNQAIITDETNDLSVAPOEKVHKSILPNSLGVSRLOLQRTFFPAVSV 1077  
QY 798 NGHDQONPDLDETTARFRVETHOKTIGGNVTKPPSLVPLESOMTKKKTGK 857  
DB 1078 NGHDQONPDLDETTARFRVETHOKTIGGNVTKPPSLVPLESOMTKKKTGK 1137  
QY 858 ENSRMEENAHGIVTEVLLGRKLQHYTDSYLGFLPWEKKYFQDLDDEESLKTQAYF 917  
DB 1138 ENSRMEENAHGIVTEVLLGRKLQHYTDSYLGFLPWEKKYFQDLDDEESLKTQAYF 1197  
QY 918 TDSKNTGROLK 928  
DB 1198 TDSKNTGROLK 1208

RESULT 14

ABR61384

ID ABR61384 standard; protein; 908 AA.

XX ABR61384;

XX 01-AUG-2003 (first entry)

XX Murine GlcNAc-phosphotransferase alpha-subunit.

XX Mouse; N-acetylglucosamine-1-phosphotransferase; nephrotropic;

XX GlcNAc-phosphotransferase; phosphodiester alpha-GlcNAcase;

XX N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase;

XX enzyme replacement therapy; phosphorylated lysosomal hydrolase;

XX lysosomal storage disease; enzyme; alpha-subunit.

XX Mus musculus.

XX US6537785-B1.

XX 25-MAR-2003.

XX 10-AUG-2000; 2000US-00636077.

XX 14-SEP-1999; 99US-0153831P.

XX (GENZ-) GENZYME GLYCOBIOLOGY RES INST INC.

XX Canfield WW;

XX WPI; 2001-290356/30.

XX N-PSDB; ACC81007.

XX Novel N-acetylglucosamine-1-phosphotransferase and N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase, useful for producing phosphorylated lysosomal hydrolase for treating lysosomal storage diseases.

XX Disclosure; Page 44-47; 62pp; English.

XX The invention relates to a novel isolated human N-acetylglucosamine-1-phosphotransferase (GlcNAc-phosphotransferase) (I) and phosphodiester

CC alpha-GlcNAcase (N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase) (ii). The protein of the invention has  
 CC nephrotropic activity, and may be useful in enzyme replacement therapy. A  
 CC protein of the invention (i), (ii) is useful for preparing a  
 CC phosphorylated lysosomal hydrolase. The phosphorylated hydrolase  
 CC comprising a terminal mannose-6-phosphate, is useful for treating a  
 CC patient suffering from a lysosomal storage disease. The present sequence  
 CC is used in the exemplification of the invention  
 XX  
 SQ Sequence 908 AA;

Query Match 79.4%; Score 3894; DB 4; Length 908;  
 Best Local Similarity 79.9%; Pred. No. 7.2e-305;  
 Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;

QY 1 MLFKLLQRTYTCLSHRYGLYVCFGLGVVTVIVSAFQGEVLEWSRDQYHVLFDSDYRNI 60  
 DB 1 MLFKLLQRTYTCLSHRYGLYVCFGLGVVTVIVSAFQGEVLEWSRDQYHVLFDSDYRNI 60  
 QY 61 AGKSFQNRCLCPMPIDVVYTWNGTDLLELKEQLQVREOQEEQKAMREILGKNTTEPTK 120  
 DB 61 AGKSFQNRCLCPMPIDVVYTWNGTDLLELKEQLQVREOQEEQKAMREILGKNTTEPTK 120  
 QY 121 KSEKQLECLLTHCIVKPMVLDPALPANITLKVPSLYSFHSGASDI FNVAKPKNPSTNV 180  
 DB 121 KSEKQLECLLTHCIVKPMVLDPALPATITLKDPTLYPSFHASSDMFNVAKPKNPSTNV 180  
 QY 181 SVVVFDTTKDVEDAHAGPFGGQQTVDWRYALTDDKAPGLVLIQGLAFSLGFPPTFKET 240  
 DB 181 PVVVFDTTKDVEDAHAGPFGGQQTVDWRYALTDDKAPGLVLIQGLAFSLGFPPTFKET 240  
 QY 241 NQTKLKPEN - LSSKVKLLQLYSEASVALLKNNPKDFQELNKTOKKNTIDGKELTISP 299  
 DB 241 SOLTKLPKRAFPKIKLLRLYSEASVALLKNNPKDFQELNKTOKKNTIDGKELTISP 300  
 QY 300 AYLLMDLSAISQSKQDEDSASRPEDNEELRYLSIERHAPWVRNIFIVTNGQIISWLN 359  
 DB 301 AYLLMDLSAISQSKQDEDSASRPEDNEELRYLSIERHAPWVRNIFIVTNGQIISWLN 360  
 QY 360 LDNPRVITVHQDVFNRLSHLPFTSSPAIESHRIEGLSQKFIYLNDDVMFGKDWPPD 419  
 DB 361 LDNPRVITVHQDIFQNLSHLPFTSSPAIESHRIEGLSQKFIYLNDDVMFGKDWPPD 420  
 QY 420 FYSHSGKQKVVLTWPVNCACGPGSWIKGYCDKACNSACDWDGDCSGNSGGRYIA 479  
 DB 421 FYSHSGKQKVVLTWPVNCACGPGSWIKGYCDKACNTSPCDWDGDCSGNTAGNRFA 480  
 QY 480 GGGTGTSGVGHWPQFGGINSVSYCQGCANSWLADKFCQACNVLSGCFDAGDCGQDH 539  
 DB 481 RGGTGTGAGQHWQFGGINTSYCQGCANSWLADKFCQACNVLSGCFDAGDCGQDH 540  
 QY 540 PHELYKVILLPNQTHYIIPKGECLPYPSFAEAVKRGVEGAYSNDPIIRHASIANKWKTH 599  
 DB 541 PHELYKVILLPNQTHYVVPKGEYLSYFSFARIARKIETGYSDNPPIIRHASIANKWKTH 600  
 QY 600 LIMSGMNTIHHNLIFQNTNDEEFKMOITVEVDTEGPKLNASTAQGVENLVSPITLL 659  
 DB 601 LIMSGMNTIHHNLIFQNTNDEEFKMOITVEVDTEGPKLNASTAQGVENLVSPITLL 660  
 QY 660 PEAILFEDIPKEKFPFKSHDVNSTRAQEEVKIPLVNIISLLPKDAQLSLNLDLQLE 719  
 DB 661 PQADVPEDEVKKEKFPKIRSHDVNATGRFQEEVKIPRVNISLIPKEAQVRLSNLDLQLE 720  
 QY 720 HGDITLKYNLSKALLSFLMNSQAKIKNOAIIITDNDLSVAPQEKQVHKSILPNSL 779  
 DB 721 RGDITLKYNLSKALLSFLMNSQAKIKNOAIIITDNDLSVAPQEKQVHKSILPNSL 777  
 QY 780 GVSRERLQRLTPPVSVKVNGHQQNPPLDLETTARFRVETHQTKTIGGNVTKEKPPSLI 839  
 DB 778 AGEHRSERWTAPAEVTVKGRDHALNPPVPLETNARL - APTGLGVTVSKENISPLI 832  
 QY 840 VPLSQMTKEKIKTGKEKERNERMEENAHNGVTEVLLGRKLQHYTDSYLGELFPEKKKY 899

DB 833 VPESHLP-----KEESDRAEGNA---VPVKELVPGRRLO---QNVPGFLPWEKKY 879  
 QY 900 FQDLDEEESIKTOLAYFTDSKNTGROLK 928  
 DB 980 FQDLDEEESIKTOLAYFTDRKHTGROLK 908  
 RESULT 15  
 ADD27817  
 ID ADD27817 standard; protein; 908 AA.  
 XX  
 AC ADD27817;  
 XX  
 DT 15-JAN-2004 (first entry)  
 XX  
 DE GlcNAc-phosphotransferase associated protein #1.  
 XX  
 KW mouse; protein phosphorylation; soluble GlcNAc-phosphotransferase;  
 KW UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease; ds.  
 XX  
 OS Mus musculus.  
 XX  
 PN US2003119088-A1.  
 XX  
 PD 26-JUN-2003.  
 XX  
 PF 21-DEC-2001; 2001US-00023888.  
 XX  
 PR 21-DEC-2001; 2001US-00023888.  
 XX  
 PA (NOVA-) NOVAYME PHARM INC.  
 XX  
 XX Canfield W, Kudo M;  
 DR WPI; 2003-801323/75.  
 XX  
 PT Phosphorylating a protein for treating a patient suffering from a  
 PT lysosomal storage disease e.g. Fabry's disease by contacting the protein  
 PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated  
 PT protein.  
 XX  
 PS Disclosure; SEQ ID NO 9; 55pp; English.  
 XX  
 CC The invention relates to a method of phosphorylating a protein comprising  
 CC contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-  
 CC acetylglucosamine) and producing a phosphorylated protein. The method is  
 CC useful for treating a patient suffering from a lysosomal storage disease  
 CC e.g. Fabry's disease. The present sequence represents the amino acid  
 CC sequence of a GlcNAc-phosphotransferase associated protein.  
 XX  
 SQ Sequence 908 AA;

Query Match 79.4%; Score 3894; DB 7; Length 908;  
 Best Local Similarity 79.9%; Pred. No. 7.2e-305;  
 Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;

QY 1 MLFKLLQRTYTCLSHRYGLYVCFGLGVVTVIVSAFQGEVLEWSRDQYHVLFDSDYRNI 60  
 DB 1 MLFKLLQRTYTCLSHRYGLYVCFGLGVVTVIVSAFQGEVLEWSRDQYHVLFDSDYRNI 60  
 QY 61 AGKSFQNRCLCPMPIDVVYTWNGTDLLELKEQLQVREOQEEQKAMREILGKNTTEPTK 120  
 DB 61 AGKSFQNRCLCPMPIDVVYTWNGTDLLELKEQLQVREOQEEQKAMREILGKNTTEPTK 120  
 QY 121 KSEKQLECLLTHCIVKPMVLDPALPANITLKVPSLYSFHSGASDI FNVAKPKNPSTNV 180  
 DB 121 KSEKQLECLLTHCIVKPMVLDPALPATITLKDPTLYPSFHASSDMFNVAKPKNPSTNV 180  
 QY 181 SVVVFDTTKDVEDAHAGPFGGQQTVDWRYALTDDKAPGLVLIQGLAFSLGFPPTFKET 240  
 DB 181 PVVVFDTTKDVEDAHAGPFGGQQTVDWRYALTDDKAPGLVLIQGLAFSLGFPPTFKET 240  
 QY 241 NQTKLKPEN - LSSKVKLLQLYSEASVALLKNNPKDFQELNKTOKKNTIDGKELTISP 299

Db	241	SQUTKLPRKAPFLKTLRLLYSEASVALLKLNPKFQELNOKTKQMTIDGKELTIS	300
Qy	300	AYLLWDLAISQSKODEDISARFEDNEELRYLSRSTIERHAPWVRNFIYVNGQIPSWLN	359
Db	301	AYLLWDLAISQSKODEDASARFEDNEELRYLSRSTIERHAPWVRNFIYVNGQIPSWLN	360
Qy	360	LDNPRVTIYTHQDVFNLSHLFTFSSPAIESHIIHREGLSOKFYIYLNDDYFQKQWPPD	419
Db	361	LDNPRVTIYTHQDIFONLSHLFTFSSPAIESHIIHREGLSOKFYIYLNDDYFQKQWPPD	420
Qy	420	FYSHSGXQKYLTWTPVNCBAECPGSWIKDGYCDKACNCSACDWDGDCSGNSGGSRYA	479
Db	421	FYSHSGXQKYLTWTPVNCBAECPGSWIKDGYCDKACNCSACDWDGDCSGNSGGSRYA	480
Qy	480	GGGTCGIGVGHWPQGGGINSVSYCNOGCANSWLADKFCDOACNVLSGCFDAGDCQDH	539
Db	481	RGGGTNGIGAGQHWQGGGINTSYCNOGCANSWLADKFCDOACNVLSGCFDAGDCQDH	540
Qy	540	FHELYKVILLPNQTHYIIPKGECLPYFSAEVAKRGEVAYSNDPIIRHASIANKWKTH	599
Db	541	FHELYKVILLPNQTHYVWPKEGYLSYFSAFIARKEGTYSNDPIIRHASIANKWKTH	600
Qy	600	LIMHSGNATTHIENLTFONTNDEEFKQITVEDTBEGPKLNSTAKGVENLSVITLL	659
Db	601	LIMPGGNATTIFYNLITQANDEEFKQIAVEDTBEPALNSTTQKAYESLVSEVPL	660
Qy	660	PEABILFEDIPKEKPPFKRHDVNSTRAQEEVKIPLVNIISLLPKDAQLSNTLDLQLE	719
Db	661	QOADVPEFDVPEKEKPPKIRRHVDVNATGRFOEBVKIPRVNIISLLPKDAQVRLSNTLDLQLE	720
Qy	720	HGDIILKGYNLSKALLSFLMNSCHAKIKNCAITDETNDLSVAQEQVHKSLIPNSL	779
Db	721	RGDIILKGYNLSKALLSFLGNSLDYKIKPCA-RTDETGNLEVPQENPSHRR--PHGF	777
Qy	780	GVSELRQRLTFPAVSKVNGHDQGNPPLDLETTARFVETHTQKTIIGNVTKEKPPSLI	839
Db	778	AGEHRSERWTAPEATVTVKGRDHALNPPVLETNARL-----AQPTLGVTVSKENLSPLI	832
Qy	840	VPLESOMTKEKKITCKEKENSREMEENHIGVTEVLLGRKLOHYTDSYLGFLPWEKKY	899
Db	833	VPPESHLP-----KEEESDRAGNA---VPVKELVFGRLQ---QNYPGFLPWEKKY	879
Qy	900	FODLDBEESLTKQLAYFTDSKNTGRLK	928
Db	880	FODLDBEESLTKQLAYFTDRKHTGRLK	908

Search completed: November 21, 2004, 12:54:44  
Job time : 112.597 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run On: November 21, 2004, 12:47:18 ; Search time 27.9723 Seconds  
(without alignments)  
2200.144 Million cell updates/sec

Title: US-10-023-888-4  
Perfect score: 4907  
Sequence: 1 MLFKLQRTYTCLSHRYGL.....SLKTLAYFTDSKNTGRQLX 928

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5E COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4907	100.0	928	4	US-09-635-872A-1
2	4907	100.0	928	4	US-09-636-077A-1
3	4907	100.0	928	4	US-09-636-060C-1
4	4907	100.0	928	4	US-09-986-552-1
5	4907	100.0	928	4	US-09-636-596C-1
6	3894	79.4	908	4	US-09-635-872A-15
7	3894	79.4	908	4	US-09-636-077A-15
8	3894	79.4	908	4	US-09-636-060C-15
9	3894	79.4	908	4	US-09-986-552-15
10	3894	79.4	908	4	US-09-636-596C-15
11	503	10.3	113	4	US-09-635-872A-11
12	503	10.3	113	4	US-09-636-077A-11
13	503	10.3	113	4	US-09-636-060C-11
14	503	10.3	113	4	US-09-986-552-11
15	503	10.3	113	4	US-09-636-596C-11
16	488	9.9	391	4	US-09-270-767-42309
17	479	9.8	502	4	US-09-635-872A-13
18	479	9.8	502	4	US-09-636-077A-13
19	479	9.8	502	4	US-09-636-060C-13
20	479	9.8	502	4	US-09-986-552-13
21	479	9.8	502	4	US-09-636-596C-13
22	265	5.4	545	4	US-08-936-107A-10
23	255.5	5.2	546	4	US-08-311-731A-248
24	240	4.9	364	4	US-09-328-352-5382
25	168.5	3.4	1015	1	US-08-537-210A-1
26	168.5	3.4	1015	3	US-09-113-825-1
27	168.5	3.4	2471	1	US-08-185-432-16

28	168.5	3.4	2471	1	US-08-083-590A-19	Sequence 19, Appl
29	168.5	3.4	2471	3	US-08-532-384-19	Sequence 19, Appl
30	168.5	3.4	2471	4	US-08-899-232-1	Sequence 1, Appl
31	168.5	3.4	2471	4	US-09-121-457-1	Sequence 1, Appl
32	154.5	3.1	2321	4	US-09-230-652-2	Sequence 2, Appl
33	152	3.1	1964	3	US-09-467-997-1	Sequence 1, Appl
34	145.5	3.0	1068	1	US-08-537-210A-2	Sequence 2, Appl
35	145.5	3.0	1068	3	US-09-113-825-2	Sequence 2, Appl
36	145.5	3.0	2556	1	US-08-185-432-17	Sequence 17, Appl
37	145.5	3.0	2556	1	US-08-083-590A-20	Sequence 20, Appl
38	145.5	3.0	2556	3	US-08-532-384-20	Sequence 20, Appl
39	145.5	3.0	2556	4	US-08-899-232-2	Sequence 2, Appl
40	145.5	3.0	2556	4	US-09-121-457-2	Sequence 2, Appl
41	141	2.9	1078	1	US-08-264-534-32	Sequence 32, Appl
42	141	2.9	1078	1	US-08-083-590A-11	Sequence 11, Appl
43	141	2.9	1078	1	US-08-465-500-32	Sequence 32, Appl
44	141	2.9	1078	2	US-08-346-128-32	Sequence 32, Appl
45	141	2.9	1078	3	US-08-532-384-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1  
US-09-635-872A-1  
; Sequence 1, Application US/09635872A  
; Patent No. 6534300  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASE.  
; FILE REFERENCE: 195613US0  
; CURRENT APPLICATION NUMBER: US/09/635,872A  
; CURRENT FILING DATE: 2000-09-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 928  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-635-872A-1

Query Match		100.0%	Score 4907;	DB 4;	Length 928;
Best Local Similarity		100.0%	Pred. No. 0;		
Matches 928;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MLFKLQRTYTCLSHRYGLYVCFLGVVTVISAFQGEVWLEWSDQVHVLFDSDYRDN	60		
Db	1	MLFKLQRTYTCLSHRYGLYVCFLGVVTVISAFQGEVWLEWSDQVHVLFDSDYRDN	60		
Qy	61	AGKSFQNRICLPMPIDVYVTVWNGTDLLELLKLOQVREOMEEOQKMRILGKNTTEPTK	120		
Db	61	AGKSFQNRICLPMPIDVYVTVWNGTDLLELLKLOQVREOMEEOQKMRILGKNTTEPTK	120		
Qy	121	KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVSLSYPSFHSASDIFNVAKPKNPSTNV	180		
Db	121	KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVSLSYPSFHSASDIFNVAKPKNPSTNV	180		
Qy	181	SVVVFSTKDVDAHSGLLKGNRSQTVWRGYLTQKEVGLVMQDLAFLSGFPPTFKET	240		
Db	181	SVVVFSTKDVDAHSGLLKGNRSQTVWRGYLTQKEVGLVMQDLAFLSGFPPTFKET	240		
Qy	241	NQLKTLPNLASKVKLLQLYSEASVALLKNNPKDFQELNKQTKNNWIDGKELTISPA	300		
Db	241	NQLKTLPNLASKVKLLQLYSEASVALLKNNPKDFQELNKQTKNNWIDGKELTISPA	300		
Qy	301	YLLWDLAISQSKQDEDISASFEONEELRYSLRSTERHAPVVRNIFIVTNGQIPSWLNL	360		
Db	301	YLLWDLAISQSKQDEDISASFEONEELRYSLRSTERHAPVVRNIFIVTNGQIPSWLNL	360		
Qy	361	DNPRVITVTHQDVFRNLHLPTFSSPAIESHRIEGLSKELIYLNDDVWFCKDVPDFF	420		

Db 361 DNPRVTIVTHQDVFRNLVSHLPTFTSSPAIESHHRIEGLSQKFYLYLNDVDMFGKDVWPDFF 420  
Qy 421 YSHSGQKVYLTWPVPCAECPGSGWIKDGYCDKACNNACDWDGDCSGNSGGSRYIAG 480  
Db 421 YSHSGQKVYLTWPVPCAECPGSGWIKDGYCDKACNNACDWDGDCSGNSGGSRYIAG 480  
Qy 481 GGGTSGSIVGHPWQFGGGINSVYCNQGCANSLADKFCDOACNVLSGCGDAGDCGQDHF 540  
Db 481 GGGTSGSIVGHPWQFGGGINSVYCNQGCANSLADKFCDOACNVLSGCGDAGDCGQDHF 540  
Qy 541 HELYKVILLPNOHYIIPKGECLPYFSPFAEVAKEGVEGAYSDNPIIRHASIANKWKTTHL 600  
Db 541 HELYKVILLPNOHYIIPKGECLPYFSPFAEVAKEGVEGAYSDNPIIRHASIANKWKTTHL 600  
Qy 601 IMHSGMNATTHFNLTQNTNDEEFKQITVEVDTRREGPKLNSTAKGYENLVSPITLLP 660  
Db 601 IMHSGMNATTHFNLTQNTNDEEFKQITVEVDTRREGPKLNSTAKGYENLVSPITLLP 660  
Qy 661 EAEILFEDIPEKEFPKRDVNSTRAOEVEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720  
Db 661 EAEILFEDIPEKEFPKRDVNSTRAOEVEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720  
Qy 721 GDITLKGYNLSKALLRSFLMNSQAKIKKQAIITDETNDLSLVAPOEKQVHKSILPNSLG 780  
Db 721 GDITLKGYNLSKALLRSFLMNSQAKIKKQAIITDETNDLSLVAPOEKQVHKSILPNSLG 780  
Qy 781 VSELRQLRTPPAVSVKNGHDQGNPPDLLETTARFRVETHQKTIGGNTKEKPPSLIV 840  
Db 781 VSELRQLRTPPAVSVKNGHDQGNPPDLLETTARFRVETHQKTIGGNTKEKPPSLIV 840  
Qy 841 PLESQMTKEKKITGKEKENSMEENAEHNHIGVTEVLLGRKLQHYTDSYLGFLPWEKKYF 900  
Db 841 PLESQMTKEKKITGKEKENSMEENAEHNHIGVTEVLLGRKLQHYTDSYLGFLPWEKKYF 900  
Qy 901 QDLLDEESLKTQLAYFTDTSKNTGRQLK 928  
Db 901 QDLLDEESLKTQLAYFTDTSKNTGRQLK 928  
RESULT 2  
US-09-636-077A-1  
; Sequence 1, Application US/09636077A  
; Patent No. 6537785  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM  
; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE  
; FILE REFERENCE: 19612US0  
; CURRENT APPLICATION NUMBER: US/09/636,077A  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 928  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-636-077A-1  
Query Match 100.0%; Score 4907; DB 4; Length 928;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MLFKLLQRQTYTCLSHRYGLYVCFGLGVVVTVISAFQGEVVLWSRQYHVLVDSYRDI 60  
Db 1 MLFKLLQRQTYTCLSHRYGLYVCFGLGVVVTVISAFQGEVVLWSRQYHVLVDSYRDI 60  
Qy 61 AGKSFQNRCLPMPIDVVYTWANGTDLLELKELOQVREOMEEBQKAMREILGNNTTEPTK 120  
Db 61 AGKSFQNRCLPMPIDVVYTWANGTDLLELKELOQVREOMEEBQKAMREILGNNTTEPTK 120  
Qy 121 KSEKQLECLLTHCIVKPMVLDPALPANITLKVPSLYPSFHSASDIFNVAKPNPSTNV 180

Db 121 KSEKQLECLLTHCIVKPMVLDPALPANITLKVPSLYPSFHSASDIFNVAKPNPSTNV 180  
Qy 181 SVVVFSDTKDVEDAHSGLLKNSRQTVWRGYLTDDKEVPGVLVLMQDLAFLSGFPPTFKET 240  
Db 181 SVVVFSDTKDVEDAHSGLLKNSRQTVWRGYLTDDKEVPGVLVLMQDLAFLSGFPPTFKET 240  
Qy 241 NOLKTKLPENISSKVKLLQLLYSEASVALLKLNPKDPQELNKPQKMTTIDGKELTISPA 300  
Db 241 NOLKTKLPENISSKVKLLQLLYSEASVALLKLNPKDPQELNKPQKMTTIDGKELTISPA 300  
Qy 301 YLLWDLISAISOSKODEDISASRFEDNEELRYSLSIERHAPWVRNIFIVINGQIPSWLNL 360  
Db 301 YLLWDLISAISOSKODEDISASRFEDNEELRYSLSIERHAPWVRNIFIVINGQIPSWLNL 360  
Qy 361 DNPRVTIVTHQDVFRNLVSHLPTFTSSPAIESHHRIEGLSQKFYLYLNDVDMFGKDVWPDFF 420  
Db 361 DNPRVTIVTHQDVFRNLVSHLPTFTSSPAIESHHRIEGLSQKFYLYLNDVDMFGKDVWPDFF 420  
Qy 421 YSHSGQKVYLTWPVPCAECPGSGWIKDGYCDKACNNACDWDGDCSGNSGGSRYIAG 480  
Db 421 YSHSGQKVYLTWPVPCAECPGSGWIKDGYCDKACNNACDWDGDCSGNSGGSRYIAG 480  
Qy 481 GGGTSGSIVGHPWQFGGGINSVYCNQGCANSLADKFCDOACNVLSGCGDAGDCGQDHF 540  
Db 481 GGGTSGSIVGHPWQFGGGINSVYCNQGCANSLADKFCDOACNVLSGCGDAGDCGQDHF 540  
Qy 541 HELYKVILLPNOHYIIPKGECLPYFSPFAEVAKEGVEGAYSDNPIIRHASIANKWKTTHL 600  
Db 541 HELYKVILLPNOHYIIPKGECLPYFSPFAEVAKEGVEGAYSDNPIIRHASIANKWKTTHL 600  
Qy 601 IMHSGMNATTHFNLTQNTNDEEFKQITVEVDTRREGPKLNSTAKGYENLVSPITLLP 660  
Db 601 IMHSGMNATTHFNLTQNTNDEEFKQITVEVDTRREGPKLNSTAKGYENLVSPITLLP 660  
Qy 661 EAEILFEDIPEKEFPKRDVNSTRAOEVEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720  
Db 661 EAEILFEDIPEKEFPKRDVNSTRAOEVEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720  
Qy 721 GDITLKGYNLSKALLRSFLMNSQAKIKKQAIITDETNDLSLVAPOEKQVHKSILPNSLG 780  
Db 721 GDITLKGYNLSKALLRSFLMNSQAKIKKQAIITDETNDLSLVAPOEKQVHKSILPNSLG 780  
Qy 781 VSELRQLRTPPAVSVKNGHDQGNPPDLLETTARFRVETHQKTIGGNTKEKPPSLIV 840  
Db 781 VSELRQLRTPPAVSVKNGHDQGNPPDLLETTARFRVETHQKTIGGNTKEKPPSLIV 840  
Qy 841 PLESQMTKEKKITGKEKENSMEENAEHNHIGVTEVLLGRKLQHYTDSYLGFLPWEKKYF 900  
Db 841 PLESQMTKEKKITGKEKENSMEENAEHNHIGVTEVLLGRKLQHYTDSYLGFLPWEKKYF 900  
Qy 901 QDLLDEESLKTQLAYFTDTSKNTGRQLK 928  
Db 901 QDLLDEESLKTQLAYFTDTSKNTGRQLK 928

## RESULT 3

US-09-636-060C-1  
; Sequence 1, Application US/09636060C  
; Patent No. 6642038  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM M  
; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY  
; FILE REFERENCE: 210119USOCONT  
; CURRENT APPLICATION NUMBER: US/09/636,060C  
; CURRENT FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 928  
; TYPE: PRT  
; ORGANISM: Homo sapiens



US-09-636-060C-1

Query Match 100.0%; Score 4907; DB 4; Length 928;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFKLLQRTYTCLSHRYGLVYCFGLGVVVTIVSAFQGEVLEWSDQYHVLFDSDYRDN 60  
DB 1 MLFKLLQRTYTCLSHRYGLVYCFGLGVVVTIVSAFQGEVLEWSDQYHVLFDSDYRDN 60

QY 61 AGKSFQNRICLPMPIIDVVTWVNGTDLLELLKELQOVREOMESEQKAMREILGKNTTPTK 120  
DB 61 AGKSFQNRICLPMPIIDVVTWVNGTDLLELLKELQOVREOMESEQKAMREILGKNTTPTK 120

QY 121 KSEKOLECILLTHCIKVPMLVLDPALPANITLKDVPSPSYPSHSDIFNVAKPSPSTNV 180  
DB 121 KSEKOLECILLTHCIKVPMLVLDPALPANITLKDVPSPSYPSHSDIFNVAKPSPSTNV 180

QY 181 SVVFPDSTKDVEDAHSGLLKGNRSQRTVMRGYLTDDKEVPLVLMQDLAFSLGFPPTFKET 240  
DB 181 SVVFPDSTKDVEDAHSGLLKGNRSQRTVMRGYLTDDKEVPLVLMQDLAFSLGFPPTFKET 240

QY 241 NOLKTKLPENISSKVKLLQLYSEASVALLKLNPKDFQELNKTQKNTIDGKELTISPA 300  
DB 241 NOLKTKLPENISSKVKLLQLYSEASVALLKLNPKDFQELNKTQKNTIDGKELTISPA 300

QY 301 YLLWDLAISQSKQDEDISASRFEDNEELRYSLRISERHAPWVRNIFIVTNGQIPSWNL 360  
DB 301 YLLWDLAISQSKQDEDISASRFEDNEELRYSLRISERHAPWVRNIFIVTNGQIPSWNL 360

QY 361 DNPRTVITVHQDVFRNLSHLPTFSSPAIESHIEHIEGLSQKFIYLNDDVDFGQVWDDF 420  
DB 361 DNPRTVITVHQDVFRNLSHLPTFSSPAIESHIEHIEGLSQKFIYLNDDVDFGQVWDDF 420

QY 421 YSHSGOKVYLTWVPNCAEGCPGSGWIKDGYCDKACNNSACDWDGDCSGSGSRYTAG 480  
DB 421 YSHSGOKVYLTWVPNCAEGCPGSGWIKDGYCDKACNNSACDWDGDCSGSGSRYTAG 480

QY 481 GGGTSGISGVHPWQFGGGINSVSYCNQGCANSWLADKFCDOACNVLSGDFDAGCGQDHF 540  
DB 481 GGGTSGISGVHPWQFGGGINSVSYCNQGCANSWLADKFCDOACNVLSGDFDAGCGQDHF 540

QY 541 HELYKVILLPNQTHYIIPKGECLPYFSAEVAKRGVAGSDNPIIRHASIANKWKTIHL 600  
DB 541 HELYKVILLPNQTHYIIPKGECLPYFSAEVAKRGVAGSDNPIIRHASIANKWKTIHL 600

QY 601 IMHSGMNATTIHFNLTFQNTNDEEFKQITVEVDTRGPKLNSTAQKGYENLVSPITLLP 660  
DB 601 IMHSGMNATTIHFNLTFQNTNDEEFKQITVEVDTRGPKLNSTAQKGYENLVSPITLLP 660

QY 661 EAEILFEDIPEKRPFPKFRHDVNSTRAQEEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720  
DB 661 EAEILFEDIPEKRPFPKFRHDVNSTRAQEEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720

QY 721 GDTILKGNLSKALLRSFLMNSOHAKIKNOAIIITDETNDLSLVAPOEKQVHKSILPNSLG 780  
DB 721 GDTILKGNLSKALLRSFLMNSOHAKIKNOAIIITDETNDLSLVAPOEKQVHKSILPNSLG 780

QY 781 VSERLQRLTTPAVSVKNGHDQGNPPDLDTTARFRVETHQKTIIGNVTKKPKPSLI 840  
DB 781 VSERLQRLTTPAVSVKNGHDQGNPPDLDTTARFRVETHQKTIIGNVTKKPKPSLI 840

QY 841 PLESQMTKEKIKTKEKENSMEENAEHNGVTEVLLGRKLQHTDYSILGFLPWEKKYF 900  
DB 841 PLESQMTKEKIKTKEKENSMEENAEHNGVTEVLLGRKLQHTDYSILGFLPWEKKYF 900

QY 901 QDLADEESLKTQLAYFTDSKNTGRQLK 928  
DB 901 QDLADEESLKTQLAYFTDSKNTGRQLK 928

RESULT 4

US-09-986-552-1

; Sequence 1, Application US/09986552  
; Patent No. 6670165  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASE;  
; FILE REFERENCE: 215089US77DIV  
; CURRENT APPLICATION NUMBER: US/09/986,552  
; CURRENT FILING DATE: 2001-11-09  
; PRIOR APPLICATION NUMBER: 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 928  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-986-552-1

Query Match 100.0%; Score 4907; DB 4; Length 928;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFKLLQRTYTCLSHRYGLVYCFGLGVVVTIVSAFQGEVLEWSDQYHVLFDSDYRDN 60  
DB 1 MLFKLLQRTYTCLSHRYGLVYCFGLGVVVTIVSAFQGEVLEWSDQYHVLFDSDYRDN 60

QY 61 AGKSFQNRICLPMPIIDVVTWVNGTDLLELLKELQOVREOMESEQKAMREILGKNTTPTK 120  
DB 61 AGKSFQNRICLPMPIIDVVTWVNGTDLLELLKELQOVREOMESEQKAMREILGKNTTPTK 120

QY 121 KSEKOLECILLTHCIKVPMLVLDPALPANITLKDVPSPSYPSHSDIFNVAKPSPSTNV 180  
DB 121 KSEKOLECILLTHCIKVPMLVLDPALPANITLKDVPSPSYPSHSDIFNVAKPSPSTNV 180

QY 181 SVVFPDSTKDVEDAHSGLLKGNRSQRTVMRGYLTDDKEVPLVLMQDLAFSLGFPPTFKET 240  
DB 181 SVVFPDSTKDVEDAHSGLLKGNRSQRTVMRGYLTDDKEVPLVLMQDLAFSLGFPPTFKET 240

QY 241 NOLKTKLPENISSKVKLLQLYSEASVALLKLNPKDFQELNKTQKNTIDGKELTISPA 300  
DB 241 NOLKTKLPENISSKVKLLQLYSEASVALLKLNPKDFQELNKTQKNTIDGKELTISPA 300

QY 301 YLLWDLAISQSKQDEDISASRFEDNEELRYSLRISERHAPWVRNIFIVTNGQIPSWNL 360  
DB 301 YLLWDLAISQSKQDEDISASRFEDNEELRYSLRISERHAPWVRNIFIVTNGQIPSWNL 360

QY 361 DNPRTVITVHQDVFRNLSHLPTFSSPAIESHIEHIEGLSQKFIYLNDDVDFGQVWDDF 420  
DB 361 DNPRTVITVHQDVFRNLSHLPTFSSPAIESHIEHIEGLSQKFIYLNDDVDFGQVWDDF 420

QY 421 YSHSGOKVYLTWVPNCAEGCPGSGWIKDGYCDKACNNSACDWDGDCSGSGSRYTAG 480  
DB 421 YSHSGOKVYLTWVPNCAEGCPGSGWIKDGYCDKACNNSACDWDGDCSGSGSRYTAG 480

QY 481 GGGTSGISGVHPWQFGGGINSVSYCNQGCANSWLADKFCDOACNVLSGDFDAGCGQDHF 540  
DB 481 GGGTSGISGVHPWQFGGGINSVSYCNQGCANSWLADKFCDOACNVLSGDFDAGCGQDHF 540

QY 541 HELYKVILLPNQTHYIIPKGECLPYFSAEVAKRGVAGSDNPIIRHASIANKWKTIHL 600  
DB 541 HELYKVILLPNQTHYIIPKGECLPYFSAEVAKRGVAGSDNPIIRHASIANKWKTIHL 600

QY 601 IMHSGMNATTIHFNLTFQNTNDEEFKQITVEVDTRGPKLNSTAQKGYENLVSPITLLP 660  
DB 601 IMHSGMNATTIHFNLTFQNTNDEEFKQITVEVDTRGPKLNSTAQKGYENLVSPITLLP 660

QY 661 EAEILFEDIPEKRPFPKFRHDVNSTRAQEEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720  
DB 661 EAEILFEDIPEKRPFPKFRHDVNSTRAQEEVKIPLVNISSLPKDAQLSLNTLDLQLEH 720

QY 721 GDTILKGNLSKALLRSFLMNSOHAKIKNOAIIITDETNDLSLVAPOEKQVHKSILPNSLG 780

Db 721 GDITLKGYNLSKALLRSFLNSOAHAKIKNOAIITDETNDLSVAPOKQVHKSILPNSLG 780  
Qy 781 VSERLQRLTTPAVSVKYNHGDQGNPPLDLETTARFRVETHQKTIGGNVTKEKPPSLIV 840  
Db 781 VSERLQRLTTPAVSVKYNHGDQGNPPLDLETTARFRVETHQKTIGGNVTKEKPPSLIV 840  
Qy 841 PLESQMTKEKITGKEKENSEMEENAHGIVTEVLLGRKLOHYTDSYLGFLPWEKKKYP 900  
Db 841 PLESQMTKEKITGKEKENSEMEENAHGIVTEVLLGRKLOHYTDSYLGFLPWEKKKYP 900  
Qy 901 QDLDDEESLKTQLAYFTDSKNTGRQLK 928  
Db 901 QDLDDEESLKTQLAYFTDSKNTGRQLK 928  
RESULT 5  
US-09-636-596C-1  
; Sequence 1, Application US/09636596C  
; Patent No. 6770468  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM  
; TITLE OF INVENTION: PHOSPHODIESTER ALPHA-GLCNase OF THE LYSOSOMAL TARGETING PATHWAY  
; FILE REFERENCE: 10929-0001-77  
; CURRENT APPLICATION NUMBER: US/09/636,596C  
; CURRENT FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 1  
; LENGTH: 928  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-636-596C-1  
Query Match 100.0%; Score 4907; DB 4; Length 928;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MLFKLLQRTYTCLSHRVGLVVCFLGVVTVTSVAFQGEVLEWSDQYHVLFDSDYRNI 60  
Db 1 MLFKLLQRTYTCLSHRVGLVVCFLGVVTVTSVAFQGEVLEWSDQYHVLFDSDYRNI 60  
Qy 61 AGKSFQNRCLCLPMPIDVVYTVWNGTDLLELLKELQVREMEEQKAMREILGKNTTEPTK 120  
Db 61 AGKSFQNRCLCLPMPIDVVYTVWNGTDLLELLKELQVREMEEQKAMREILGKNTTEPTK 120  
Qy 121 KSEKQLECLLTHCTIKVPMVLDPALPANITLKVPSLYPSFHSASDIPNVAKPKNSTNV 180  
Db 121 KSEKQLECLLTHCTIKVPMVLDPALPANITLKVPSLYPSFHSASDIPNVAKPKNSTNV 180  
Qy 181 SVVVPDSTKDYEDAHSGLLKNSRQTVWRGVLTTDKVPGVLVMDLAFSLGPPPTFKET 240  
Db 181 SVVVPDSTKDYEDAHSGLLKNSRQTVWRGVLTTDKVPGVLVMDLAFSLGPPPTFKET 240  
Qy 241 NQTKLPENISSKVKLLQLYSEASVALLKLNPNKDFQELNKQTKQNTDGLKELTISP 300  
Db 241 NQTKLPENISSKVKLLQLYSEASVALLKLNPNKDFQELNKQTKQNTDGLKELTISP 300  
Qy 301 YLLWDLISAISOKODEDASRFDNEELRSLRSIERHAPVWNIIVNGIIPSWNL 360  
Db 301 YLLWDLISAISOKODEDASRFDNEELRSLRSIERHAPVWNIIVNGIIPSWNL 360  
Qy 361 DNPRTVITHQDVFNLSHLPTFSSPAIESHIHRIEGLSQFIYLVNDVDFKDPDPP 420  
Db 361 DNPRTVITHQDVFNLSHLPTFSSPAIESHIHRIEGLSQFIYLVNDVDFKDPDPP 420  
Qy 421 YSHSGQKQVLTWVPNCAEGCPGSIWKDGYCDKACNNSACDWDGDCSGNSGGSRYIAG 480  
Db 421 YSHSGQKQVLTWVPNCAEGCPGSIWKDGYCDKACNNSACDWDGDCSGNSGGSRYIAG 480  
Qy 481 GGGTSGIGVGHFWQFGGINSVSYCNGCANSLADKFCDOACNVLSGCGFDAGCGQDHF 540

Db 481 GGGTSGIGVGHFWQFGGINSVSYCNGCANSLADKFCDOACNVLSGCGFDAGCGQDHF 540  
Qy 541 HELYKIVILLPNQTHYIIPKGECLPYFSPAEVAKGVEGAYSDNPRIIRHASIANKWKTIHL 600  
Db 541 HELYKIVILLPNQTHYIIPKGECLPYFSPAEVAKGVEGAYSDNPRIIRHASIANKWKTIHL 600  
Qy 601 IMHSGMAATTIHFNLTFTONTNDESFQKQITVEVDTRGPKLNSTAQKGYENLVSPIITLLP 660  
Db 601 IMHSGMAATTIHFNLTFTONTNDESFQKQITVEVDTRGPKLNSTAQKGYENLVSPIITLLP 660  
Qy 661 EAEILFEDIPKEKPPKFRHVDVNSTRAQSEVKIPLVNSLLPKDAQLSLNTLDLQLEH 720  
Db 661 EAEILFEDIPKEKPPKFRHVDVNSTRAQSEVKIPLVNSLLPKDAQLSLNTLDLQLEH 720  
Qy 721 GDITLKGYNLSKALLRSFLNSOAHAKIKNOAIITDETNDLSVAPOKQVHKSILPNSLG 780  
Db 721 GDITLKGYNLSKALLRSFLNSOAHAKIKNOAIITDETNDLSVAPOKQVHKSILPNSLG 780  
Qy 781 VSERLQRLTTPAVSVKYNHGDQGNPPLDLETTARFRVETHQKTIGGNVTKEKPPSLIV 840  
Db 781 VSERLQRLTTPAVSVKYNHGDQGNPPLDLETTARFRVETHQKTIGGNVTKEKPPSLIV 840  
Qy 841 PLESQMTKEKITGKEKENSEMEENAHGIVTEVLLGRKLOHYTDSYLGFLPWEKKKYP 900  
Db 841 PLESQMTKEKITGKEKENSEMEENAHGIVTEVLLGRKLOHYTDSYLGFLPWEKKKYP 900  
Qy 901 QDLDDEESLKTQLAYFTDSKNTGRQLK 928  
Db 901 QDLDDEESLKTQLAYFTDSKNTGRQLK 928  
RESULT 6  
US-09-635-872A-15  
; Sequence 15, Application US/09635872A  
; Patent No. 6534300  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 195613USO  
; CURRENT APPLICATION NUMBER: US/09/635,872A  
; CURRENT FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 15  
; LENGTH: 908  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-635-872A-15  
Query Match 79.4%; Score 3894; DB 4; Length 908;  
Best Local Similarity 79.9%; Pred. No. 0;  
Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;  
Qy 1 MLFKLLQRTYTCLSHRVGLVVCFLGVVTVTSVAFQGEVLEWSDQYHVLFDSDYRNI 60  
Db 1 MLFKLLQRTYTCLSHRVGLVVCFLGVVTVTSVAFQGEVLEWSDQYHVLFDSDYRNI 60  
Qy 61 AGKSFQNRCLCLPMPIDVVYTVWNGTDLLELLKELQVREMEEQKAMREILGKNTTEPTK 120  
Db 61 AGKSFQNRCLCLPMPIDVVYTVWNGTDLLELLKELQVREMEEQKAMREILGKNTTEPTK 120  
Qy 121 KSEKQLECLLTHCTIKVPMVLDPALPANITLKVPSLYPSFHSASDIPNVAKPKNSTNV 180  
Db 121 KSEKQLECLLTHCTIKVPMVLDPALPANITLKVPSLYPSFHSASDIPNVAKPKNSTNV 180  
Qy 181 SVVVPDSTKDYEDAHSGLLKNSRQTVWRGVLTTDKVPGVLVMDLAFSLGPPPTFKET 240  
Db 181 SVVVPDSTKDYEDAHSGLLKNSRQTVWRGVLTTDKVPGVLVMDLAFSLGPPPTFKET 240  
Qy 241 NQTKLPENISSKVKLLQLYSEASVALLKLNPNKDFQELNKQTKQNTDGLKELTISP 299

```
Db 241 SOLKTKLPRKAPFKIKILRLYSEASVALLKNNPKGFQELNKQTKKNTIDGKELTISP 300
QY 300 AYLLMDLSAISOSKODEIDISARFEDNEELRYSLRSIERHAPWVRNIFIVTNGQIPSWLN 359
Db 301 AYLLMDLSAISOSKODEIDISARFEDNEELRYSLRSIERHAPWVRNIFIVTNGQIPSWLN 360
QY 360 LNPRTVITVTHQDVPRNLSHLPSTSSPAIESHRIEGLSOKFYLDNDVDFGKDVPPDD 419
Db 361 LNPRTVITVTHQDVPRNLSHLPSTSSPAIESHRIEGLSOKFYLDNDVDFGKDVPPDD 420
QY 420 FYSHSGKQKYLTPVPNCABGCGPSWIKDGYCDKACNNSACDMDGDCSGNSGGRYIA 479
Db 421 FYSHSGKQKYLTPVPNCABGCGPSWIKDGYCDKACNNSPCDMDGDCSGNTAGNFVA 480
QY 480 GGGTGNIGAGQHWFGGGINTISYCNQGCANSWLADKFCDOACNVLSGCGDAGCGQDH 539
Db 481 RGGTGNIGAGQHWFGGGINTISYCNQGCANSWLADKFCDOACNVLSGCGDAGCGQDH 540
QY 540 FHELYKVLLPNQTHYIIPKGECLPYFSFAEVAKEGVEGAYSDNPIIRHASIANKWKTIIH 599
Db 541 FHELYKVLLPNQTHYIIPKGECLPYFSFAEVAKEGVEGAYSDNPIIRHASIANKWKTIIH 600
QY 600 LIMHSGMNTATTHFNLTQNTNDEEFKQITVEVDTRGPKLNSTAKQGYENLVSPITLL 659
Db 601 LIMPGMNATTIYFNLTLQNDANDEEFKQIAVEVDTRGPKLNSTAKQGYENLVSPITLL 660
QY 660 PEABTLFEDIPEKRPFKRHDVNSTRAAEVKIPLVNISSLPKDAQLSLNLDLQLE 719
Db 661 PQADVPFEDVPKRPFKRHDVNSTRAAEVKIPLVNISSLPKDAQLSLNLDLQLE 720
QY 720 HGDITLKGYNLSKALLRSFLMNSQAKIKNOAIITDETNDLSLVAPOEKQVHKSIILPNSL 779
Db 721 RGDITLKGYNLSKALLRSFLMNSQAKIKNOAIITDETNDLSLVAPOEKQVHKSIILPNSL 780
QY 780 GVSELRQRLTTPPAVSVKNGHDQGNPPDLDETTARFVRVETHQTKTIGNVTKEKPPSLI 839
Db 778 AGEHRSERWTAPAEITVTKGRDHALNPPVPVLETNARL-----AQPTLGVTVSKENLSPLI 832
QY 840 VPLESQMTKEKKITGKEKENSMEENHIGVTEVLLGRKLQHYTDSYLGFLPWEKKY 899
Db 833 VPESHLP-----KEESDRAEGNA---VPVKELVGRQLQ---QNYPGFLPWEKKY 879
QY 900 FQDLDDEESLKTQLAYFTDSKNTGRQLK 928
Db 880 FQDLDDEESLKTQLAYFTDRKHTGRQLK 908
```

## RESULT 7

```
US-09-636-077A-15
; Sequence 15, Application US/09636077A
; Patent No. 6537785
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE
; FILE REFERENCE: 195612US0
; CURRENT APPLICATION NUMBER: US/09/636,077A
; CURRENT FILING DATE: 2000-08-10
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 908
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-636-077A-15
```

Query Match 79.4%; Score 3894; DB 4; Length 908;

Best Local Similarity 79.9%; Pred No. 0;

Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;

QY 1 MLFKLLQRTYTCLSHRYGLYVCFGLGVVVVTIVSAFQFGEVVLWSRQYHVLFDSDYRDN 60

```
Db 1 MLFKLLQRTYTCLSHRYGLYVCFGLGVVVVTIVSAFQFGEVVLWSRQYHVLFDSDYRDN 60
QY 61 AGKSFQNRCLPMPIDVYVTVWNGTDLELLKELQQVREOMESBQKAMREILGNNTTEPTK 120
Db 61 AGKSFQNRCLPMPIDVYVTVWNGTDLELLKELQQVREOMESBQKAMREILGNNTTEPTK 120
QY 121 KSEKQLECLLTHCHIKVPMLVLDPALPANITLKDVPSPLYSPHSASDIFNVAKPKPNSTNV 180
Db 121 KSEKQLECLLTHCHIKVPMLVLDPALPANITLKDVPSPLYSPHSASDIFNVAKPKPNSTNV 180
QY 181 SVVVFDTKDVEDAHAGLLKGNRSQRTVWRGYLTDTKEVPLGLVLMQDLAFLSGPPPTFKET 240
Db 181 PVVVFDTKDVEDAHAGLLKGNRSQRTVWRGYLTDTKEVPLGLVLMQDLAFLSGPPPTFKET 240
QY 241 NQKTKLKPEN-LSSKVKLLQLYSEASVALLKNNPKGFQELNKQTKKNTIDGKELTISP 299
Db 241 SQUKTKLPRKAPFKIKILRLYSEASVALLKNNPKGFQELNKQTKKNTIDGKELTISP 300
QY 300 AYLLMDLSAISOSKODEIDISARFEDNEELRYSLRSIERHAPWVRNIFIVTNGQIPSWLN 359
Db 301 AYLLMDLSAISOSKODEIDISARFEDNEELRYSLRSIERHAPWVRNIFIVTNGQIPSWLN 360
QY 360 LNPRTVITVTHQDVPRNLSHLPSTSSPAIESHRIEGLSOKFYLDNDVDFGKDVPPDD 419
Db 361 LNPRTVITVTHQDVPRNLSHLPSTSSPAIESHRIEGLSOKFYLDNDVDFGKDVPPDD 420
QY 420 FYSHSGKQKYLTPVPNCABGCGPSWIKDGYCDKACNNSACDMDGDCSGNSGGRYIA 479
Db 421 FYSHSGKQKYLTPVPNCABGCGPSWIKDGYCDKACNNSPCDMDGDCSGNTAGNFVA 480
QY 480 GGGTGNIGAGQHWFGGGINTISYCNQGCANSWLADKFCDOACNVLSGCGDAGCGQDH 539
Db 481 RGGTGNIGAGQHWFGGGINTISYCNQGCANSWLADKFCDOACNVLSGCGDAGCGQDH 540
QY 540 FHELYKVLLPNQTHYIIPKGECLPYFSFAEVAKEGVEGAYSDNPIIRHASIANKWKTIIH 599
Db 541 FHELYKVLLPNQTHYIIPKGECLPYFSFAEVAKEGVEGAYSDNPIIRHASIANKWKTIIH 600
QY 600 LIMHSGMNTATTHFNLTQNTNDEEFKQITVEVDTRGPKLNSTAKQGYENLVSPITLL 659
Db 601 LIMPGMNATTIYFNLTLQNDANDEEFKQIAVEVDTRGPKLNSTAKQGYENLVSPITLL 660
QY 660 PEABTLFEDIPEKRPFKRHDVNSTRAAEVKIPLVNISSLPKDAQLSLNLDLQLE 719
Db 661 PQADVPFEDVPKRPFKRHDVNSTRAAEVKIPLVNISSLPKDAQLSLNLDLQLE 720
QY 720 HGDITLKGYNLSKALLRSFLMNSQAKIKNOAIITDETNDLSLVAPOEKQVHKSIILPNSL 779
Db 721 RGDITLKGYNLSKALLRSFLMNSQAKIKNOAIITDETNDLSLVAPOEKQVHKSIILPNSL 780
QY 780 GVSELRQRLTTPPAVSVKNGHDQGNPPDLDETTARFVRVETHQTKTIGNVTKEKPPSLI 839
Db 778 AGEHRSERWTAPAEITVTKGRDHALNPPVPVLETNARL-----AQPTLGVTVSKENLSPLI 832
QY 840 VPLESQMTKEKKITGKEKENSMEENHIGVTEVLLGRKLQHYTDSYLGFLPWEKKY 899
Db 833 VPESHLP-----KEESDRAEGNA---VPVKELVGRQLQ---QNYPGFLPWEKKY 879
QY 900 FQDLDDEESLKTQLAYFTDSKNTGRQLK 928
Db 880 FQDLDDEESLKTQLAYFTDRKHTGRQLK 908
```

## RESULT 8

```
US-09-636-060C-15
; Sequence 15, Application US/09636060C
; Patent No. 6642038
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 21019US0CONT
; CURRENT APPLICATION NUMBER: US/09/636,060C
```

```
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 15
; LENGTH: 908
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-636-060C-15

Query Match      79.4%; Score 3894; DB 4; Length 908;
Best Local Similarity 79.9%; Pred. No. 0;
Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;

QY 1 MLFKLLQRTYTCLSHRYGLVYCFGLGVVVTVISAFQFGEVLEWSRDQYHVLFDSDYRDN 60
Db 1 MLFKLLQRTYTCLSHRYGLVYCFGLGVVVTVISAFQFGEVLEWSRDQYHVLFDSDYRDN 60
QY 61 AGKSFQNLCLPMPIDVVYTWVNGTDLLELLKELQVREMEBEQKAMREILGKNTTEPTK 120
Db 61 AGKSFQNLCLPMPIDVVYTWVNGTDLLELLKELQVREMEBEQKAMREILGKNTTEPTK 120
QY 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPISLPYSPHSDIFNVAKPNPSTNV 180
Db 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPISLPYSPHSDIFNVAKPNPSTNV 180
QY 181 SVVVPDSTKDVDEAHSGLLKGNRSQTVWRGYLTDDKEVPGVLVLMQDLAFLSGPPTPKET 240
Db 181 SVVVPDSTKDVDEAHSGLLKGNRSQTVWRGYLTDDKEVPGVLVLMQDLAFLSGPPTPKET 240
QY 241 NQKTKLPEN-LSSKVKLLQLYSEASVALLKNNPKDFQELNKOTKKNMTIDGKELTISP 299
Db 241 NQKTKLPEN-LSSKVKLLQLYSEASVALLKNNPKDFQELNKOTKKNMTIDGKELTISP 299
QY 300 AYLLWDLAISQSKQDEDISASRFEDNEELRYSLRSIERHAPVWRNIPIVINGQIPSWLN 359
Db 300 AYLLWDLAISQSKQDEDISASRFEDNEELRYSLRSIERHAPVWRNIPIVINGQIPSWLN 359
QY 360 LONPRVITVTHQDVPRLNLSHLPSTSSPAIESHRIEGLSKFYIYNDVDFGKDWDPDD 419
Db 360 LONPRVITVTHQDVPRLNLSHLPSTSSPAIESHRIEGLSKFYIYNDVDFGKDWDPDD 419
QY 420 FVSHSGKQKYLTPVNPNCABGCPGSKWIKDGYCDKACNNSACDWDGDCSGNSGSRVIA 479
Db 420 FVSHSGKQKYLTPVNPNCABGCPGSKWIKDGYCDKACNNSACDWDGDCSGNSGSRVIA 479
QY 480 GGGTGSIGVGHMPWQFGGGINSVYCNQGCANSWLADKFCDOACNVLSGCGDAGCGGDH 539
Db 480 GGGTGSIGVGHMPWQFGGGINSVYCNQGCANSWLADKFCDOACNVLSGCGDAGCGGDH 539
QY 481 RGGGTGNIGAGQHWQFGGINTISYCNQGCANSWLADKFCDOACNVLSGCGDAGCGGDH 540
Db 481 RGGGTGNIGAGQHWQFGGINTISYCNQGCANSWLADKFCDOACNVLSGCGDAGCGGDH 540
QY 540 FHELYKVILLNTHQYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWKTLH 599
Db 540 FHELYKVILLNTHQYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWKTLH 599
QY 541 FHELYKVILLNTHQYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWKTLH 600
Db 541 FHELYKVILLNTHQYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWKTLH 600

; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 15
; LENGTH: 908
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-636-060C-15

Query Match      79.4%; Score 3894; DB 4; Length 908;
Best Local Similarity 79.9%; Pred. No. 0;
Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;

QY 1 MLFKLLQRTYTCLSHRYGLVYCFGLGVVVTVISAFQFGEVLEWSRDQYHVLFDSDYRDN 60
Db 1 MLFKLLQRTYTCLSHRYGLVYCFGLGVVVTVISAFQFGEVLEWSRDQYHVLFDSDYRDN 60
QY 61 AGKSFQNLCLPMPIDVVYTWVNGTDLLELLKELQVREMEBEQKAMREILGKNTTEPTK 120
Db 61 AGKSFQNLCLPMPIDVVYTWVNGTDLLELLKELQVREMEBEQKAMREILGKNTTEPTK 120
QY 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPISLPYSPHSDIFNVAKPNPSTNV 180
Db 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPISLPYSPHSDIFNVAKPNPSTNV 180
QY 181 SVVVPDSTKDVDEAHSGLLKGNRSQTVWRGYLTDDKEVPGVLVLMQDLAFLSGPPTPKET 240
Db 181 SVVVPDSTKDVDEAHSGLLKGNRSQTVWRGYLTDDKEVPGVLVLMQDLAFLSGPPTPKET 240
QY 241 NQKTKLPEN-LSSKVKLLQLYSEASVALLKNNPKDFQELNKOTKKNMTIDGKELTISP 299
Db 241 NQKTKLPEN-LSSKVKLLQLYSEASVALLKNNPKDFQELNKOTKKNMTIDGKELTISP 299
QY 300 AYLLWDLAISQSKQDEDISASRFEDNEELRYSLRSIERHAPVWRNIPIVINGQIPSWLN 359
Db 300 AYLLWDLAISQSKQDEDISASRFEDNEELRYSLRSIERHAPVWRNIPIVINGQIPSWLN 359
QY 360 LONPRVITVTHQDVPRLNLSHLPSTSSPAIESHRIEGLSKFYIYNDVDFGKDWDPDD 419
Db 360 LONPRVITVTHQDVPRLNLSHLPSTSSPAIESHRIEGLSKFYIYNDVDFGKDWDPDD 419
QY 420 FVSHSGKQKYLTPVNPNCABGCPGSKWIKDGYCDKACNNSACDWDGDCSGNSGSRVIA 479
Db 420 FVSHSGKQKYLTPVNPNCABGCPGSKWIKDGYCDKACNNSACDWDGDCSGNSGSRVIA 479
QY 480 GGGTGSIGVGHMPWQFGGGINSVYCNQGCANSWLADKFCDOACNVLSGCGDAGCGGDH 539
Db 480 GGGTGSIGVGHMPWQFGGGINSVYCNQGCANSWLADKFCDOACNVLSGCGDAGCGGDH 539
QY 481 RGGGTGNIGAGQHWQFGGINTISYCNQGCANSWLADKFCDOACNVLSGCGDAGCGGDH 540
Db 481 RGGGTGNIGAGQHWQFGGINTISYCNQGCANSWLADKFCDOACNVLSGCGDAGCGGDH 540
QY 540 FHELYKVILLNTHQYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWKTLH 599
Db 540 FHELYKVILLNTHQYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWKTLH 599
QY 541 FHELYKVILLNTHQYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWKTLH 600
Db 541 FHELYKVILLNTHQYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWKTLH 600

; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 15
; LENGTH: 908
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-636-060C-15

Query Match      79.4%; Score 3894; DB 4; Length 908;
Best Local Similarity 79.9%; Pred. No. 0;
Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;

QY 1 MLFKLLQRTYTCLSHRYGLVYCFGLGVVVTVISAFQFGEVLEWSRDQYHVLFDSDYRDN 60
Db 1 MLFKLLQRTYTCLSHRYGLVYCFGLGVVVTVISAFQFGEVLEWSRDQYHVLFDSDYRDN 60
QY 61 AGKSFQNLCLPMPIDVVYTWVNGTDLLELLKELQVREMEBEQKAMREILGKNTTEPTK 120
Db 61 AGKSFQNLCLPMPIDVVYTWVNGTDLLELLKELQVREMEBEQKAMREILGKNTTEPTK 120
QY 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPISLPYSPHSDIFNVAKPNPSTNV 180
Db 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPISLPYSPHSDIFNVAKPNPSTNV 180
QY 181 SVVVPDSTKDVDEAHSGLLKGNRSQTVWRGYLTDDKEVPGVLVLMQDLAFLSGPPTPKET 240
Db 181 SVVVPDSTKDVDEAHSGLLKGNRSQTVWRGYLTDDKEVPGVLVLMQDLAFLSGPPTPKET 240
QY 241 NQKTKLPEN-LSSKVKLLQLYSEASVALLKNNPKDFQELNKOTKKNMTIDGKELTISP 299
Db 241 NQKTKLPEN-LSSKVKLLQLYSEASVALLKNNPKDFQELNKOTKKNMTIDGKELTISP 299
QY 300 AYLLWDLAISQSKQDEDISASRFEDNEELRYSLRSIERHAPVWRNIPIVINGQIPSWLN 359
Db 300 AYLLWDLAISQSKQDEDISASRFEDNEELRYSLRSIERHAPVWRNIPIVINGQIPSWLN 359
QY 360 LONPRVITVTHQDVPRLNLSHLPSTSSPAIESHRIEGLSKFYIYNDVDFGKDWDPDD 419
Db 360 LONPRVITVTHQDVPRLNLSHLPSTSSPAIESHRIEGLSKFYIYNDVDFGKDWDPDD 419
QY 420 FVSHSGKQKYLTPVNPNCABGCPGSKWIKDGYCDKACNNSACDWDGDCSGNSGSRVIA 479
Db 420 FVSHSGKQKYLTPVNPNCABGCPGSKWIKDGYCDKACNNSACDWDGDCSGNSGSRVIA 479
QY 480 GGGTGSIGVGHMPWQFGGGINSVYCNQGCANSWLADKFCDOACNVLSGCGDAGCGGDH 539
Db 480 GGGTGSIGVGHMPWQFGGGINSVYCNQGCANSWLADKFCDOACNVLSGCGDAGCGGDH 539
QY 481 RGGGTGNIGAGQHWQFGGINTISYCNQGCANSWLADKFCDOACNVLSGCGDAGCGGDH 540
Db 481 RGGGTGNIGAGQHWQFGGINTISYCNQGCANSWLADKFCDOACNVLSGCGDAGCGGDH 540
QY 540 FHELYKVILLNTHQYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWKTLH 599
Db 540 FHELYKVILLNTHQYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWKTLH 599
QY 541 FHELYKVILLNTHQYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWKTLH 600
Db 541 FHELYKVILLNTHQYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWKTLH 600
```

QY 600 LIMHSGMNAITTHFNLTFTQNTNDBEFKMOITVEVDTRGPKLNSTAKQYENLVSPITLL 659  
DB 601 LIMPGGMNATIIYFNLTQNDANDEEFKIOIAVEVDTRAPKLNSTTQKAYESLVSPVPL 660  
QY 660 PEAEILFEDIPEKRFPPKFRHDVNSTRAQEEVKIPLVNIISLPPKDAQLSLNTLDLQLE 719  
DB 661 PQADVPEFDPKRFPPKFRHDVNSTRAQEEVKIPLVNIISLPPKDAQLSLNTLDLQLE 720  
QY 720 HGDITLKGYNLSKALLSFLMNSQHAKIKNOAIIITDETNDLSLVAPOEKQVHKSLPNSL 779  
DB 721 RGDITLKGYNLSKALLSFLMNSQHAKIKNOAIIITDETNDLSLVAPOEKQVHKSLPNSL 777  
QY 780 GVSERLQRLTTPAVSVKYNHGDQGNPPDLLETTARFRVETHTQKTGGNVTKKPPSLI 839  
DB 778 AGEHRSERWTAPAEVTTVKGRDHALNPPVPLETNARL-----AQPTLGTVSKENLSPLI 832  
QY 840 VPLESOMTKKKTGKEKENSMEENAHIGHVTEVLLGRKLOHYTDSYLGFLPWEKKY 899  
DB 833 VPPEHLP-----KEESDRAEGNA---VPVKELVPGRRLO---QNYPGFLPWEKKY 879  
QY 900 FQDILLDEESLKTQLAYFTDSKNTGRQLK 928  
DB 880 FQDILLDEESLKTQLAYFTDRKHTGRQLK 908

## RESULT 10

US-09-636-596C-15

; Sequence 15, Application US/09636596C

; Patent No. 6770468

; GENERAL INFORMATION:

; APPLICANT: CANFIELD, WILLIAM

; TITLE OF INVENTION: PHOSPHODIESTER ALPHA-GLUCANase OF THE LYSOSOMAL TARGETING PATHWAY

; FILE REFERENCE: 10929-0001-77

; CURRENT APPLICATION NUMBER: US/09/636.596C

; CURRENT FILING DATE: 2000-08-10

; PRIOR APPLICATION NUMBER: 60/153,831

; PRIOR FILING DATE: 1999-09-14

; NUMBER OF SEQ ID NOS: 52

; SOFTWARE: Patent in version 3.0

; SEQ ID NO 15

; LENGTH: 908

; TYPE: PRT

; ORGANISM: Mus musculus

US-09-636-596C-15

Query Match 79.4%; Score 3894; DB 4; Length 908;

Best Local Similarity 79.9%; Pred. No. 0;

Matches 742; Conservative 61; Mismatches 104; Indels 22; Gaps 7;

QY 1 MLFKLQROVTYCLSHRYGLVVCFLGVVTVTSAFQGEVVLKSRDOYHVLFDSDYDNI 60  
DB 1 MLFKLQROVTYCLSHRYGLVVCFLGVVTVTSAFQGEVVLKSRDOYHVLFDSDYDNI 60  
QY 61 AGKSFQNRCLPMPIDVVYTWNGTDLLELLEKQVREOQMEERELGKNTTEPTK 120  
DB 61 AGKSFQNRCLPMPIDVVYTWNGTDLLELLEKQVREOQMEERELGKNTTEPTK 120  
QY 121 KSEKQLECLLTHCIKVPMLVLDLPALPANITLKVPSLYPSFHSASDIFNVAKPKNPSTNV 180  
DB 121 KSEKQLECLLTHCIKVPMLVLDLPALPANITLKVPSLYPSFHSASDIFNVAKPKNPSTNV 180  
QY 181 SVWVFDSTKDVEDAHSGLLKGNSTQTVMRGYLTDDKVPGLVLMODLAFLSGPPPTFKET 240  
DB 181 PVVVFDTKDVEDAHAGFPKGGQGTQVWRAVLTDDKAPGLVLIQGLAFLSGPPPTFKET 240  
QY 241 NQTKTKLPEN--LSKVKLLQLYSEASVALLKNNPKDFQELNKTKQKMTIDGKELTISP 299  
DB 241 SOLTKTKLPKAPFLKIKLLRLYSEASVALLKNNPKGFQELNKTKQKMTIDGKELTISP 300  
QY 300 AYLLWDLISLQSKODEDIIASRFEDNEELRYSLRSTERRAPVNRNIFIVTNGQIPSWLN 359  
DB 301 AYLLWDLISLQSKODEDIIASRFEDNEELRYSLRSTERRAPVNRNIFIVTNGQIPSWLN 360

QY 360 LDMPRVTIYTHODVFRNLSHLPTFFSPATESHRIIEGLSOKFIYLNDDVFMFGKWPPD 419  
DB 361 LDMPRVTIYTHODVFRNLSHLPTFFSPATESHRIIEGLSOKFIYLNDDVFMFGKWPPD 420  
QY 420 FYSHSGQGVYLTWVPNCAEGCPGSMIXDGYCDKACNNSACDWDGCGDSCGSGSGGSRVIA 479  
DB 421 FYSHSGQGVYLTWVPNCAEGCPGSMIXDGYCDKACNNSPCDWDGCGNCGTAGNRFVA 480  
QY 480 GGGGTGSGIVGHPWQGGGGINSVSYCNCANSLWADKFCDOACNVLSGFGDAGCGQDH 539  
DB 481 RGGGTGSGIVGHPWQGGGGINSVSYCNCANSLWADKFCDOACNVLSGFGDAGCGQDH 540  
QY 540 FHELYKVILLPQOYTHYIIIPKGECLPYFSAEVAKEGVEGAGYSDNPRIIRHASIANKWKTH 599  
DB 541 FHELYKVILLPQOYTHYIIIPKGECLPYFSAEVAKEGVEGAGYSDNPRIIRHASIANKWKTH 600  
QY 600 LIMHSGMNAITTHFNLTFTQNTNDBEFKMOITVEVDTRGPKLNSTAKQYENLVSPITLL 659  
DB 601 LIMPGGMNATIIYFNLTQNDANDEEFKIOIAVEVDTRAPKLNSTTQKAYESLVSPVPL 660  
QY 660 PEAEILFEDIPEKRFPPKFRHDVNSTRAQEEVKIPLVNIISLPPKDAQLSLNTLDLQLE 719  
DB 661 PQADVPEFDPKRFPPKFRHDVNSTRAQEEVKIPLVNIISLPPKDAQLSLNTLDLQLE 720  
QY 720 HGDITLKGYNLSKALLSFLMNSQHAKIKNOAIIITDETNDLSLVAPOEKQVHKSLPNSL 779  
DB 721 RGDITLKGYNLSKALLSFLMNSQHAKIKNOAIIITDETNDLSLVAPOEKQVHKSLPNSL 777  
QY 780 GVSERLQRLTTPAVSVKYNHGDQGNPPDLLETTARFRVETHTQKTGGNVTKKPPSLI 839  
DB 778 AGEHRSERWTAPAEVTTVKGRDHALNPPVPLETNARL-----AQPTLGTVSKENLSPLI 832  
QY 840 VPLESOMTKKKTGKEKENSMEENAHIGHVTEVLLGRKLOHYTDSYLGFLPWEKKY 899  
DB 833 VPPEHLP-----KEESDRAEGNA---VPVKELVPGRRLO---QNYPGFLPWEKKY 879  
QY 900 FQDILLDEESLKTQLAYFTDSKNTGRQLK 928  
DB 880 FQDILLDEESLKTQLAYFTDRKHTGRQLK 908

## RESULT 11

US-09-635-872A-11

; Sequence 11, Application US/09635872A

; Patent No. 6534300

; GENERAL INFORMATION:

; APPLICANT: CANFIELD, WILLIAM

; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES

; FILE REFERENCE: 195613US0

; CURRENT APPLICATION NUMBER: US/09/635,872A

; CURRENT FILING DATE: 2000-08-10

; PRIOR APPLICATION NUMBER: 60/153,831

; PRIOR FILING DATE: 1999-09-14

; NUMBER OF SEQ ID NOS: 52

; SOFTWARE: Patent in version 3.0

; SEQ ID NO 11

; LENGTH: 113

; TYPE: PRT

; ORGANISM: Rattus rattus

US-09-635-872A-11

Query Match 10.3%; Score 503; DB 4; Length 113;

Best Local Similarity 92.5%; Pred. No. 1,2e-36;

Matches 99; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 233 FPPTFKETNQLTKLPENLSKVKLLQLYSEASVALLKNNPKDFQELNKTKQKMTIDG 292  
DB 1 FPPTFKETNQLTKLPENLSKVKLLQLYSEASVALLKNNPKGFPELNKTKQKMTIDG 60  
QY 293 KELTISPAVLLWDLISLQSKODEDIIASRFEDNEELRYSLRSTERR 339  
DB 61 KELTISPAVLLWDLISLQSKODEDIIASRFEDNEELRYSLRSTERR 107

## TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES

US-09-636-077A-11  
; Sequence 11, Application US/09636077A  
; Patent No. 6537785  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM  
; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE  
; FILE REFERENCE: 195612USO  
; CURRENT APPLICATION NUMBER: US/09/636,077A  
; CURRENT FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 11  
; LENGTH: 113  
; TYPE: PRT  
; ORGANISM: Rattus rattus  
US-09-986-552-11

Query Match 10.3%; Score 503; DB 4; Length 113;  
Best Local Similarity 92.5%; Pred. No. 1.2e-36;  
Matches 99; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 233 PPPTFKETNQLTKLPENLSSKVLLQLYSEASVALLKLNPKDFQELNKQTKKNTIDG 292  
Db 1 PPPTFKETSQTKLPENLSSKIKLLQLYSEASVALLKLNPKGFFPELNKQTKKNSISG 60  
Qy 293 KELTISPAYLLWDLISAISQSKQDEDISASRPEDNEELRYSLRSIERH 339  
Db 61 KELAISPAYLLWDLISAISQSKQDESDVSASRPEDNEELRYSLRSIERH 107

## RESULT 15

US-09-636-596C-11  
; Sequence 11, Application US/09636596C  
; Patent No. 6770468  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM  
; TITLE OF INVENTION: PHOSPHODIESTER ALPHA-CLONACASE OF THE LYSOSOMAL TARGETING PATHWAY  
; FILE REFERENCE: 10929-0001-77  
; CURRENT APPLICATION NUMBER: US/09/636,596C  
; CURRENT FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 11  
; LENGTH: 113  
; TYPE: PRT  
; ORGANISM: Rattus rattus  
US-09-636-596C-11

Query Match 10.3%; Score 503; DB 4; Length 113;  
Best Local Similarity 92.5%; Pred. No. 1.2e-36;  
Matches 99; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 233 PPPTFKETNQLTKLPENLSSKVLLQLYSEASVALLKLNPKDFQELNKQTKKNTIDG 292  
Db 1 PPPTFKETSQTKLPENLSSKIKLLQLYSEASVALLKLNPKGFFPELNKQTKKNSISG 60  
Qy 293 KELTISPAYLLWDLISAISQSKQDEDISASRPEDNEELRYSLRSIERH 339  
Db 61 KELAISPAYLLWDLISAISQSKQDESDVSASRPEDNEELRYSLRSIERH 107

Search completed: November 21, 2004, 13:03:15  
Job time : 30.9723 secs

US-09-636-077A-11  
; Sequence 11, Application US/09636077A  
; Patent No. 6537785  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM  
; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE  
; FILE REFERENCE: 195612USO  
; CURRENT APPLICATION NUMBER: US/09/636,077A  
; CURRENT FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 11  
; LENGTH: 113  
; TYPE: PRT  
; ORGANISM: Rattus rattus  
US-09-636-077A-11

Query Match 10.3%; Score 503; DB 4; Length 113;  
Best Local Similarity 92.5%; Pred. No. 1.2e-36;  
Matches 99; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 233 PPPTFKETNQLTKLPENLSSKVLLQLYSEASVALLKLNPKDFQELNKQTKKNTIDG 292  
Db 1 PPPTFKETSQTKLPENLSSKIKLLQLYSEASVALLKLNPKGFFPELNKQTKKNSISG 60  
Qy 293 KELTISPAYLLWDLISAISQSKQDEDISASRPEDNEELRYSLRSIERH 339  
Db 61 KELAISPAYLLWDLISAISQSKQDESDVSASRPEDNEELRYSLRSIERH 107

## RESULT 13

US-09-636-060C-11  
; Sequence 11, Application US/09636060C  
; Patent No. 6642038  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM M  
; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY  
; FILE REFERENCE: 210119USOCNT  
; CURRENT APPLICATION NUMBER: US/09/636,060C  
; CURRENT FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 11  
; LENGTH: 113  
; TYPE: PRT  
; ORGANISM: Rattus rattus  
US-09-636-060C-11

Query Match 10.3%; Score 503; DB 4; Length 113;  
Best Local Similarity 92.5%; Pred. No. 1.2e-36;  
Matches 99; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 233 PPPTFKETNQLTKLPENLSSKVLLQLYSEASVALLKLNPKDFQELNKQTKKNTIDG 292  
Db 1 PPPTFKETSQTKLPENLSSKIKLLQLYSEASVALLKLNPKGFFPELNKQTKKNSISG 60  
Qy 293 KELTISPAYLLWDLISAISQSKQDEDISASRPEDNEELRYSLRSIERH 339  
Db 61 KELAISPAYLLWDLISAISQSKQDESDVSASRPEDNEELRYSLRSIERH 107

## RESULT 14

US-09-986-552-11  
; Sequence 11, Application US/09986552  
; Patent No. 6670165  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2004, 12:56:09 ; Search time 100.171 Seconds  
(without alignments)  
3280.691 Million cell updates/sec

Title: US-10-023-888-4

Perfect score: 4907

Sequence: 1 MLFKLLQRTYTCLSHRYGL.....SLKQLAYFTDKNTGRQLK 928

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1570615 seqs, 354127592 residues

Total number of hits satisfying chosen parameters: 1570615

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4907	100.0	928	9	US-09-895-072-1
2	4907	100.0	928	9	US-09-986-552-1
3	4907	100.0	928	14	US-10-023-888-4
4	4907	100.0	928	14	US-10-023-889-4
5	4907	100.0	928	14	US-10-023-890-4
6	4907	100.0	928	14	US-10-024-197-4
7	4907	100.0	928	14	US-10-023-894-4
8	4907	100.0	928	14	US-10-306-686-1
9	4630	94.4	1199	14	US-10-023-888-2
10	4630	94.4	1199	14	US-10-023-889-2
11	4630	94.4	1199	14	US-10-023-890-2
12	4630	94.4	1199	14	US-10-024-197-2
13	4630	94.4	1199	14	US-10-023-894-2

```

14 4473 91.2 847 14 US-10-120-801-88 Sequence 88, Appl
15 4468 91.1 846 14 US-10-094-749-2392 Sequence 2392, Ap
16 4391 89.5 1459 14 US-10-120-801-26 Sequence 26, Appl
17 3894 79.4 908 9 US-09-895-072-15 Sequence 15, Appl
18 3894 79.4 908 9 US-09-986-552-15 Sequence 15, Appl
19 3894 79.4 908 14 US-10-023-888-9 Sequence 9, Appl
20 3894 79.4 908 14 US-10-023-889-9 Sequence 9, Appl
21 3894 79.4 908 14 US-10-023-890-9 Sequence 9, Appl
22 3894 79.4 908 14 US-10-024-197-9 Sequence 9, Appl
23 3894 79.4 908 14 US-10-023-894-9 Sequence 9, Appl
24 3894 79.4 908 14 US-10-306-686-15 Sequence 15, Appl
25 2021 41.2 663 9 US-10-120-801-87 Sequence 87, Appl
26 1900 38.7 367 9 US-09-864-761-42893 Sequence 42893, A
27 1900 38.7 367 14 US-10-023-886-32927 Sequence 32927, A
28 1292 26.3 384 14 US-10-139-794-173 Sequence 173, Appl
29 734.5 15.0 384 14 US-10-120-801-89 Sequence 89, Appl
30 503 10.3 113 9 US-09-895-072-11 Sequence 11, Appl
31 503 10.3 113 9 US-09-986-552-11 Sequence 11, Appl
32 503 10.3 113 14 US-10-023-888-14 Sequence 14, Appl
33 503 10.3 113 14 US-10-023-889-14 Sequence 14, Appl
34 503 10.3 113 14 US-10-023-890-14 Sequence 14, Appl
35 503 10.3 113 14 US-10-024-197-14 Sequence 14, Appl
36 503 10.3 113 14 US-10-023-894-14 Sequence 14, Appl
37 503 10.3 113 14 US-10-306-686-11 Sequence 11, Appl
38 488 9.9 652 14 US-10-120-801-91 Sequence 91, Appl
39 479 9.8 502 9 US-09-895-072-13 Sequence 13, Appl
40 479 9.8 502 9 US-09-986-552-13 Sequence 13, Appl
41 479 9.8 502 14 US-10-023-888-16 Sequence 16, Appl
42 479 9.8 502 14 US-10-023-889-16 Sequence 16, Appl
43 479 9.8 502 14 US-10-023-890-16 Sequence 16, Appl
44 479 9.8 502 14 US-10-024-197-16 Sequence 16, Appl
45 479 9.8 502 14 US-10-023-894-16 Sequence 16, Appl

```

#### ALIGNMENTS

```

RESULT 1
US-09-895-072-1
; Sequence 1. Application US/09895072
; Patent No. US20020025550A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM M
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 210119USOCONT
; CURRENT APPLICATION NUMBER: US/09/895,072
; CURRENT FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; PRIOR APPLICATION NUMBER: US 09/535,872
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-895-072-1

```

```

Query Match 100.0%; Score 4907; DB 9; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFKLLQRTYTCLSHRYGLVVCFLGVVVTVISAFQFGEVLEWSDQYHVLFDSDYRNI 60
DB 1 MLFKLLQRTYTCLSHRYGLVVCFLGVVVTVISAFQFGEVLEWSDQYHVLFDSDYRNI 60
QY 61 AGKSFQNRCLCPMPIDVYTWNGTDLLELQVREQMEEQKAMREILGKNTTEPTK 120
DB 61 AGKSFQNRCLCPMPIDVYTWNGTDLLELQVREQMEEQKAMREILGKNTTEPTK 120
QY 121 KSEQLECLLTHCIKVPMLVLDPALNITLKVPSLPFSFASDIFNVAKPNPSTNV 180

```

Db 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPISLYPSFHSASDIFNVAKPKNPSTNV 180  
Qy 181 SVVVFSTKDVEDAHSGLLKGNRSQTVWGYLTITDKVEPGLVLMQDLAFSLGFPPTFKET 240  
Db 181 SVVVFSTKDVEDAHSGLLKGNRSQTVWGYLTITDKVEPGLVLMQDLAFSLGFPPTFKET 240  
Qy 241 NQKTKLPENLSKVKLLQLYSEASVALLKNNPKDFQELNKOTKKNMTIDGKELTISPA 300  
Db 241 NQKTKLPENLSKVKLLQLYSEASVALLKNNPKDFQELNKOTKKNMTIDGKELTISPA 300  
Qy 301 YLLWDLAISQSQKODEDISASRFEDNEELRYSLRSIERHAPVWRNIFIVTNGQIPSWLNL 360  
Db 301 YLLWDLAISQSQKODEDISASRFEDNEELRYSLRSIERHAPVWRNIFIVTNGQIPSWLNL 360  
Qy 361 DNPRVTIVTHQDVFNRLSHLPTFSSPAIESHIHRIEGLSQKFYLLNDDVWFGKDVWDDF 420  
Db 361 DNPRVTIVTHQDVFNRLSHLPTFSSPAIESHIHRIEGLSQKFYLLNDDVWFGKDVWDDF 420  
Qy 421 YSHSGOKVYLTWVPNCAEGCGSWIKDGYCDKACNNSACDWDGDCSGNSGGSRYIAG 480  
Db 421 YSHSGOKVYLTWVPNCAEGCGSWIKDGYCDKACNNSACDWDGDCSGNSGGSRYIAG 480  
Qy 481 GGGTSGIVGHPWQFGGINSVSYCNOGCANSWLADKFCDOACNVLSGDFDAGDCQDHF 540  
Db 481 GGGTSGIVGHPWQFGGINSVSYCNOGCANSWLADKFCDOACNVLSGDFDAGDCQDHF 540  
Qy 541 HELYKVILLPQOYTHIIPKGECLPYFSPAFAVAKGVGAYSDNPIIRHASIANKWKTIHL 600  
Db 541 HELYKVILLPQOYTHIIPKGECLPYFSPAFAVAKGVGAYSDNPIIRHASIANKWKTIHL 600  
Qy 601 IMHSGMNATTIHFNLTFTQNTDEEFKQITVEVDTREGPKLNSTAKGYENLVSPITLLP 660  
Db 601 IMHSGMNATTIHFNLTFTQNTDEEFKQITVEVDTREGPKLNSTAKGYENLVSPITLLP 660  
Qy 661 EAEILFEDIPEKRPFKRHDVNSTRAQOEYKIPLVNLSLAPKDAQLSNTLDLQLEH 720  
Db 661 EAEILFEDIPEKRPFKRHDVNSTRAQOEYKIPLVNLSLAPKDAQLSNTLDLQLEH 720  
Qy 721 GDITLKGYNLSKALLRSFLMNSQHAKIKNQAIITDETNDLSLVAPOEKQVHKSILPNSLG 780  
Db 721 GDITLKGYNLSKALLRSFLMNSQHAKIKNQAIITDETNDLSLVAPOEKQVHKSILPNSLG 780  
Qy 781 VSELRQLRTPPAVSVKNGHDQGNPDLLETTARFRVETHQKTIGGNVTKKPPSLIV 840  
Db 781 VSELRQLRTPPAVSVKNGHDQGNPDLLETTARFRVETHQKTIGGNVTKKPPSLIV 840  
Qy 841 PLESQMTKEKKTIGKEKENSMEENAEENHIGVTEVLLGRKLQHYTDSYLGFLPWEKKKYF 900  
Db 841 PLESQMTKEKKTIGKEKENSMEENAEENHIGVTEVLLGRKLQHYTDSYLGFLPWEKKKYF 900  
Qy 901 QDILLDEESLKTQLAYFTDSKNTGRQLK 928  
Db 901 QDILLDEESLKTQLAYFTDSKNTGRQLK 928

RESULT 2  
US-09-986-552-1  
; Sequence 1, Application US/09986552  
; Patent No. US20020150981A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 21508905/77DIV  
; CURRENT APPLICATION NUMBER: US/09/986.552  
; CURRENT FILING DATE: 2001-11-09  
; PRIOR APPLICATION NUMBER: 09/635, 872  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153, 831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 1  
; LENGTH: 928

; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-986-552-1  
Query Match 100.0%; Score 4907; DB 9; Length 928;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MLFKLLQRYTCLSHRYGLYVCFGLVVVTIVSAFOFGEVVLWESRDQHVLFDSYRDN 60  
Db 1 MLFKLLQRYTCLSHRYGLYVCFGLVVVTIVSAFOFGEVVLWESRDQHVLFDSYRDN 60  
Qy 61 AGKSFQNRCLPMPIDVVYTWNGTDLLELKELOVRBQEBEQQKAMREILGNITTEPTK 120  
Db 61 AGKSFQNRCLPMPIDVVYTWNGTDLLELKELOVRBQEBEQQKAMREILGNITTEPTK 120  
Qy 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPISLYPSFHSASDIFNVAKPKNPSTNV 180  
Db 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPISLYPSFHSASDIFNVAKPKNPSTNV 180  
Qy 181 SVVVFSTKDVEDAHSGLLKGNRSQTVWGYLTITDKVEPGLVLMQDLAFSLGFPPTFKET 240  
Db 181 SVVVFSTKDVEDAHSGLLKGNRSQTVWGYLTITDKVEPGLVLMQDLAFSLGFPPTFKET 240  
Qy 241 NQKTKLPENLSKVKLLQLYSEASVALLKNNPKDFQELNKOTKKNMTIDGKELTISPA 300  
Db 241 NQKTKLPENLSKVKLLQLYSEASVALLKNNPKDFQELNKOTKKNMTIDGKELTISPA 300  
Qy 301 YLLWDLAISQSQKODEDISASRFEDNEELRYSLRSIERHAPVWRNIFIVTNGQIPSWLNL 360  
Db 301 YLLWDLAISQSQKODEDISASRFEDNEELRYSLRSIERHAPVWRNIFIVTNGQIPSWLNL 360  
Qy 361 DNPRVTIVTHQDVFNRLSHLPTFSSPAIESHIHRIEGLSQKFYLLNDDVWFGKDVWDDF 420  
Db 361 DNPRVTIVTHQDVFNRLSHLPTFSSPAIESHIHRIEGLSQKFYLLNDDVWFGKDVWDDF 420  
Qy 421 YSHSGOKVYLTWVPNCAEGCGSWIKDGYCDKACNNSACDWDGDCSGNSGGSRYIAG 480  
Db 421 YSHSGOKVYLTWVPNCAEGCGSWIKDGYCDKACNNSACDWDGDCSGNSGGSRYIAG 480  
Qy 481 GGGTSGIVGHPWQFGGINSVSYCNOGCANSWLADKFCDOACNVLSGDFDAGDCQDHF 540  
Db 481 GGGTSGIVGHPWQFGGINSVSYCNOGCANSWLADKFCDOACNVLSGDFDAGDCQDHF 540  
Qy 541 HELYKVILLPQOYTHIIPKGECLPYFSPAFAVAKGVGAYSDNPIIRHASIANKWKTIHL 600  
Db 541 HELYKVILLPQOYTHIIPKGECLPYFSPAFAVAKGVGAYSDNPIIRHASIANKWKTIHL 600  
Qy 601 IMHSGMNATTIHFNLTFTQNTDEEFKQITVEVDTREGPKLNSTAKGYENLVSPITLLP 660  
Db 601 IMHSGMNATTIHFNLTFTQNTDEEFKQITVEVDTREGPKLNSTAKGYENLVSPITLLP 660  
Qy 661 EAEILFEDIPEKRPFKRHDVNSTRAQOEYKIPLVNLSLAPKDAQLSNTLDLQLEH 720  
Db 661 EAEILFEDIPEKRPFKRHDVNSTRAQOEYKIPLVNLSLAPKDAQLSNTLDLQLEH 720  
Qy 721 GDITLKGYNLSKALLRSFLMNSQHAKIKNQAIITDETNDLSLVAPOEKQVHKSILPNSLG 780  
Db 721 GDITLKGYNLSKALLRSFLMNSQHAKIKNQAIITDETNDLSLVAPOEKQVHKSILPNSLG 780  
Qy 781 VSELRQLRTPPAVSVKNGHDQGNPDLLETTARFRVETHQKTIGGNVTKKPPSLIV 840  
Db 781 VSELRQLRTPPAVSVKNGHDQGNPDLLETTARFRVETHQKTIGGNVTKKPPSLIV 840  
Qy 841 PLESQMTKEKKTIGKEKENSMEENAEENHIGVTEVLLGRKLQHYTDSYLGFLPWEKKKYF 900  
Db 841 PLESQMTKEKKTIGKEKENSMEENAEENHIGVTEVLLGRKLQHYTDSYLGFLPWEKKKYF 900  
Qy 901 QDILLDEESLKTQLAYFTDSKNTGRQLK 928  
Db 901 QDILLDEESLKTQLAYFTDSKNTGRQLK 928

## RESULT 3

US-10-023-888-4  
; Sequence 4, Application US/10023888  
; Publication No. US20030119088A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: SOLUBLE GLCNAC PHOSPHOTRANSFERASE  
; FILE REFERENCE: 203515US77  
; CURRENT APPLICATION NUMBER: US/10/023,888  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 928  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-023-888-4

Query Match 100.0%; Score 4907; DB 14; Length 928;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFKLLQRTYTCLSHRYGLVCFGLGVVVTIVSAFQFGEVLEWSDQYHVLFDSDYRNI 60  
DB 1 MLFKLLQRTYTCLSHRYGLVCFGLGVVVTIVSAFQFGEVLEWSDQYHVLFDSDYRNI 60

QY 61 AGKSFQNRCLCLPMPIDVVVYTWNGTDLLELLEKQVREOMEEQKAMREILGKNTTEPTK 120  
DB 61 AGKSFQNRCLCLPMPIDVVVYTWNGTDLLELLEKQVREOMEEQKAMREILGKNTTEPTK 120

QY 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSPSPHSASDIFNVAKPKNPSTNV 180  
DB 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSPSPHSASDIFNVAKPKNPSTNV 180

QY 181 SVVVPDSTKDVEDAHSGLLKGNRSQTVWRGYLTDDKEVGLVLMQDLAFSGPPTFKET 240  
DB 181 SVVVPDSTKDVEDAHSGLLKGNRSQTVWRGYLTDDKEVGLVLMQDLAFSGPPTFKET 240

QY 241 NQKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNTIDGKELTISPA 300  
DB 241 NQKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNTIDGKELTISPA 300

QY 301 YLLWDLAISQSKODEDTSASRFEDNEELRYSLRSIERHAPWVRNI FIVTNGQIPSWNL 360  
DB 301 YLLWDLAISQSKODEDTSASRFEDNEELRYSLRSIERHAPWVRNI FIVTNGQIPSWNL 360

QY 361 DNPRVTIVTHQDVFRNLSHLPTFSSPAIESHRIEGLSQKFIYLNDDVMFGKDVWPDF 420  
DB 361 DNPRVTIVTHQDVFRNLSHLPTFSSPAIESHRIEGLSQKFIYLNDDVMFGKDVWPDF 420

QY 421 YSHSGKQKYLTPVPNCACGPGSWIKDGYCDKACNNSACDWDGDCSGNSGGSRVYAG 480  
DB 421 YSHSGKQKYLTPVPNCACGPGSWIKDGYCDKACNNSACDWDGDCSGNSGGSRVYAG 480

QY 481 GGGTSGISGVHPWQFGGINSVSYCNQGCANSWLADKFCDOACNVLSGCGFDAGCGQDHF 540  
DB 481 GGGTSGISGVHPWQFGGINSVSYCNQGCANSWLADKFCDOACNVLSGCGFDAGCGQDHF 540

QY 541 HELYKVLIPNQTHTYIIPKGCLEPVEGFAEVAKEGVSAGSDNPIIRHASTANKWKTILH 600  
DB 541 HELYKVLIPNQTHTYIIPKGCLEPVEGFAEVAKEGVSAGSDNPIIRHASTANKWKTILH 600

QY 601 IMHSGMNATTIHFNLTFCNTNDEBFKQITVEVDTRREGPKLNSTAKGYENLVSPITLLP 660  
DB 601 IMHSGMNATTIHFNLTFCNTNDEBFKQITVEVDTRREGPKLNSTAKGYENLVSPITLLP 660

QY 661 EAEILFEDIPEKEFPFKRHDVNSTRAQBEVKIPLVNTSLLPKDAQLSINTLDLQLEH 720  
DB 661 EAEILFEDIPEKEFPFKRHDVNSTRAQBEVKIPLVNTSLLPKDAQLSINTLDLQLEH 720

QY 721 GDITLKGYNLSKALLSFLMNSQAKIKNOAIIITDETNDLSVAPQEKVHKSILPNSLG 780  
DB 721 GDITLKGYNLSKALLSFLMNSQAKIKNOAIIITDETNDLSVAPQEKVHKSILPNSLG 780

## RESULT 4

US-10-023-889-4  
; Sequence 4, Application US/10023889  
; Publication No. US20030124652A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS OF PRODUCING HIGH MANNOSE GLYCOPROTEINS IN COMPLEX CARB;  
; FILE REFERENCE: 203512US77  
; CURRENT APPLICATION NUMBER: US/10/023,889  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 928  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-023-889-4

Query Match 100.0%; Score 4907; DB 14; Length 928;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFKLLQRTYTCLSHRYGLVCFGLGVVVTIVSAFQFGEVLEWSDQYHVLFDSDYRNI 60  
DB 1 MLFKLLQRTYTCLSHRYGLVCFGLGVVVTIVSAFQFGEVLEWSDQYHVLFDSDYRNI 60

QY 61 AGKSFQNRCLCLPMPIDVVVYTWNGTDLLELLEKQVREOMEEQKAMREILGKNTTEPTK 120  
DB 61 AGKSFQNRCLCLPMPIDVVVYTWNGTDLLELLEKQVREOMEEQKAMREILGKNTTEPTK 120

QY 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSPSPHSASDIFNVAKPKNPSTNV 180  
DB 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSPSPHSASDIFNVAKPKNPSTNV 180

QY 181 SVVVPDSTKDVEDAHSGLLKGNRSQTVWRGYLTDDKEVGLVLMQDLAFSGPPTFKET 240  
DB 181 SVVVPDSTKDVEDAHSGLLKGNRSQTVWRGYLTDDKEVGLVLMQDLAFSGPPTFKET 240

QY 241 NQKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNTIDGKELTISPA 300  
DB 241 NQKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNTIDGKELTISPA 300

QY 301 YLLWDLAISQSKODEDTSASRFEDNEELRYSLRSIERHAPWVRNI FIVTNGQIPSWNL 360  
DB 301 YLLWDLAISQSKODEDTSASRFEDNEELRYSLRSIERHAPWVRNI FIVTNGQIPSWNL 360

QY 361 DNPRVTIVTHQDVFRNLSHLPTFSSPAIESHRIEGLSQKFIYLNDDVMFGKDVWPDF 420  
DB 361 DNPRVTIVTHQDVFRNLSHLPTFSSPAIESHRIEGLSQKFIYLNDDVMFGKDVWPDF 420

QY 421 YSHSGKQKYLTPVPNCACGPGSWIKDGYCDKACNNSACDWDGDCSGNSGGSRVYAG 480  
DB 421 YSHSGKQKYLTPVPNCACGPGSWIKDGYCDKACNNSACDWDGDCSGNSGGSRVYAG 480

QY 481 GGGTSGISGVHPWQFGGINSVSYCNQGCANSWLADKFCDOACNVLSGCGFDAGCGQDHF 540  
DB 481 GGGTSGISGVHPWQFGGINSVSYCNQGCANSWLADKFCDOACNVLSGCGFDAGCGQDHF 540

QY 541 HELYKVILLPNQTHYIIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWKTIHL 600  
DB 541 HELYKVILLPNQTHYIIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWKTIHL 600  
QY 601 IMHSGMNAATTIHFNLITFQNTNDEEFKQIITVEVDTREGPKLNSTAKGYENLVSPITLLP 660  
DB 601 IMHSGMNAATTIHFNLITFQNTNDEEFKQIITVEVDTREGPKLNSTAKGYENLVSPITLLP 660  
QY 661 EAEILFEDIPEKEKFPKFKHVDVNSTRAQAEVVKIPLVNIISLLPKDAQLSLNTLDLQLEH 720  
DB 661 EAEILFEDIPEKEKFPKFKHVDVNSTRAQAEVVKIPLVNIISLLPKDAQLSLNTLDLQLEH 720  
QY 721 GDIITLKGYNLSKALLRSFLMNSHAKIKNOAIIITDETNDLSVAPOEKQVHKSIILPNSLG 780  
DB 721 GDIITLKGYNLSKALLRSFLMNSHAKIKNOAIIITDETNDLSVAPOEKQVHKSIILPNSLG 780  
QY 781 VSELRQRLTTPAVSVKVNKHGQGNPPDLLETTARFRVETHTKTIGGNVTKEKPPSLIV 840  
DB 781 VSELRQRLTTPAVSVKVNKHGQGNPPDLLETTARFRVETHTKTIGGNVTKEKPPSLIV 840  
QY 841 PLESQMTKEKKITGKEKENSMEENAEHNHIGVTEVLLGRKLQHYTDSYLGFLPWEKKYF 900  
DB 841 PLESQMTKEKKITGKEKENSMEENAEHNHIGVTEVLLGRKLQHYTDSYLGFLPWEKKYF 900  
QY 901 QDLLEBESLKTOLAYFTDSKNTGRQLK 928  
DB 901 QDLLEBESLKTOLAYFTDSKNTGRQLK 928  
RESULT 5  
US-10-023-890-4  
; Sequence 4, Application US/10023890  
; Publication No. US20030124653A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHOD OF PRODUCING GLYCOPROTEINS HAVING REDUCED COMPLEX CARBOHYD  
; FILE REFERENCE: 203510US77  
; CURRENT APPLICATION NUMBER: US/10/023,890  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: Patent In version 3.1  
; SEQ ID NO 4  
; LENGTH: 928  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-023-890-4  
Query Match 100.0%; Score 4907; DB 14; Length 928;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MLFKLLQRTQTYTCLSHRYGLVCFGLGVVTVIVSAFQGEVLEWSDRQYHVLFDSDYRDN 60  
DB 1 MLFKLLQRTQTYTCLSHRYGLVCFGLGVVTVIVSAFQGEVLEWSDRQYHVLFDSDYRDN 60  
QY 61 AKGSFQRLCLPMPIDVVYTWNGTDLLELLKELQVREOQEEBOKAMREILGKNTTEPTK 120  
DB 61 AKGSFQRLCLPMPIDVVYTWNGTDLLELLKELQVREOQEEBOKAMREILGKNTTEPTK 120  
QY 121 KSEKQLECLLTHCICKVPMVLDPALPANITLKVPSLYPSFHSASDIFNVAKPNSTNV 180  
DB 121 KSEKQLECLLTHCICKVPMVLDPALPANITLKVPSLYPSFHSASDIFNVAKPNSTNV 180  
QY 181 SVVVFSDTKQVEDAHSGLLKNSQTVWRGVLTTDKVPGVLMDQIAFLSGPPPTFKET 240  
DB 181 SVVVFSDTKQVEDAHSGLLKNSQTVWRGVLTTDKVPGVLMDQIAFLSGPPPTFKET 240  
QY 241 NQLXTKLPENTSSKVKLLQLYSEASVALLKLNPKDQELNKQTKGNTIDGKELTISPA 300  
DB 241 NQLXTKLPENTSSKVKLLQLYSEASVALLKLNPKDQELNKQTKGNTIDGKELTISPA 300  
QY 301 YLLWDLISAISQKODEDISASRPFEDNEELRYLSRSIERHAPVWRNIPIVINGQIPSWLNL 360

DB 301 YLLWDLISAISQKODEDISASRPFEDNEELRYLSRSIERHAPVWRNIPIVINGQIPSWLNL 360  
QY 361 DNPRVTIIVTHQDVFNLSHLPTFSSPAIESHIHIEGLSQFIYLNDDVNFVKDVPDDF 420  
DB 361 DNPRVTIIVTHQDVFNLSHLPTFSSPAIESHIHIEGLSQFIYLNDDVNFVKDVPDDF 420  
QY 421 YSHKQKQVYLITWVPNCABECPCGSWIKDGYCDKACNNSACDMDGDCSGNSGGSRYIAG 480  
DB 421 YSHKQKQVYLITWVPNCABECPCGSWIKDGYCDKACNNSACDMDGDCSGNSGGSRYIAG 480  
QY 481 GGGTGSIGVGPWPQGGGINSVSYCNOGCANSWLADKFCQACNVLSGCPDAGCGQDHF 540  
DB 481 GGGTGSIGVGPWPQGGGINSVSYCNOGCANSWLADKFCQACNVLSGCPDAGCGQDHF 540  
QY 541 HELYKVILLPNQTHYIIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWKTIHL 600  
DB 541 HELYKVILLPNQTHYIIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIANKWKTIHL 600  
QY 601 IMHSGMNAATTIHFNLITFQNTNDEEFKQIITVEVDTREGPKLNSTAKGYENLVSPITLLP 660  
DB 601 IMHSGMNAATTIHFNLITFQNTNDEEFKQIITVEVDTREGPKLNSTAKGYENLVSPITLLP 660  
QY 661 EAEILFEDIPEKEKFPKFKHVDVNSTRAQAEVVKIPLVNIISLLPKDAQLSLNTLDLQLEH 720  
DB 661 EAEILFEDIPEKEKFPKFKHVDVNSTRAQAEVVKIPLVNIISLLPKDAQLSLNTLDLQLEH 720  
QY 721 GDIITLKGYNLSKALLRSFLMNSHAKIKNOAIIITDETNDLSVAPOEKQVHKSIILPNSLG 780  
DB 721 GDIITLKGYNLSKALLRSFLMNSHAKIKNOAIIITDETNDLSVAPOEKQVHKSIILPNSLG 780  
QY 781 VSELRQRLTTPAVSVKVNKHGQGNPPDLLETTARFRVETHTKTIGGNVTKEKPPSLIV 840  
DB 781 VSELRQRLTTPAVSVKVNKHGQGNPPDLLETTARFRVETHTKTIGGNVTKEKPPSLIV 840  
QY 841 PLESQMTKEKKITGKEKENSMEENAEHNHIGVTEVLLGRKLQHYTDSYLGFLPWEKKYF 900  
DB 841 PLESQMTKEKKITGKEKENSMEENAEHNHIGVTEVLLGRKLQHYTDSYLGFLPWEKKYF 900  
QY 901 QDLLEBESLKTOLAYFTDSKNTGRQLK 928  
DB 901 QDLLEBESLKTOLAYFTDSKNTGRQLK 928  
RESULT 6  
US-10-024-197-4  
; Sequence 4, Application US/10024197  
; Publication No. US20030133924A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: HIGHLY PHOSPHORYLATED ACID BETA-GLUCOCEREEROSIDASE AND METHODS C  
; FILE REFERENCE: 209794US0  
; CURRENT APPLICATION NUMBER: US/10/024,197  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: Patent In version 3.1  
; SEQ ID NO 4  
; LENGTH: 928  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-024-197-4  
Query Match 100.0%; Score 4907; DB 14; Length 928;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MLFKLLQRTQTYTCLSHRYGLVCFGLGVVTVIVSAFQGEVLEWSDRQYHVLFDSDYRDN 60  
DB 1 MLFKLLQRTQTYTCLSHRYGLVCFGLGVVTVIVSAFQGEVLEWSDRQYHVLFDSDYRDN 60  
QY 61 AKGSFQRLCLPMPIDVVYTWNGTDLLELLKELQVREOQEEBOKAMREILGKNTTEPTK 120

```
Db 61 AGKSFQNRCLCLPMPIDVVYTWVNGTDLLELLKELQOVREQMEEEQKAMREILGNKNTTEPTK 120
Qy 121 KSEKQLECLLTHCICKVPMVLDPALPANITLKDVPSLYPSFHSASDIFNVAKPKNPSTNV 180
Db 121 KSEKQLECLLTHCICKVPMVLDPALPANITLKDVPSLYPSFHSASDIFNVAKPKNPSTNV 180
Qy 181 SVVVFSTKDVDEAHSGLLKGNRSQTVWRGVLTTDKVEPGLVLMQDLAFSLGPPPTPKET 240
Db 181 SVVVFSTKDVDEAHSGLLKGNRSQTVWRGVLTTDKVEPGLVLMQDLAFSLGPPPTPKET 240
Qy 241 NQKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKNMTIDGKELTISPA 300
Db 241 NQKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKNMTIDGKELTISPA 300
Qy 301 YLLWDLAISQKODEDISASRFEDNEELRYSLSIERHAPWVRNI FIVTNGQIPSWLNL 360
Db 301 YLLWDLAISQKODEDISASRFEDNEELRYSLSIERHAPWVRNI FIVTNGQIPSWLNL 360
Qy 361 DNPRVTIVTHQDVFRNLHLPTFSSPAIESHHRIEGLSQFIYLNDDVMFGKDVWPDF 420
Db 361 DNPRVTIVTHQDVFRNLHLPTFSSPAIESHHRIEGLSQFIYLNDDVMFGKDVWPDF 420
Qy 421 YSHSGQKQVYLTWPVNCACGPGSWIKDGYCDKACNNSACDWDGDCSGNSGSRVIAG 480
Db 421 YSHSGQKQVYLTWPVNCACGPGSWIKDGYCDKACNNSACDWDGDCSGNSGSRVIAG 480
Qy 481 GGGTSGIGVGHFWPGFQGGGINSVSYCNQGCANSLADKFCDOACNVLSGCFDAGCGGDHF 540
Db 481 GGGTSGIGVGHFWPGFQGGGINSVSYCNQGCANSLADKFCDOACNVLSGCFDAGCGGDHF 540
Qy 541 HELYKVILLPNQTHYIIPKGECLPYFSPAFAVAKRGVEGAYSDNPIIRHASIANKWKTIHL 600
Db 541 HELYKVILLPNQTHYIIPKGECLPYFSPAFAVAKRGVEGAYSDNPIIRHASIANKWKTIHL 600
Qy 601 IMHSGMNAITTHFNLTFTQNTNDEEFKQITVEVDTRGPKLNSTAOQYENLVSPITLLP 660
Db 601 IMHSGMNAITTHFNLTFTQNTNDEEFKQITVEVDTRGPKLNSTAOQYENLVSPITLLP 660
Qy 661 EAEILFEDIPKEKPPFKRHDVNSTRAQBEVKIPLVNIISLLPKDAQLSLNTLDLQLEH 720
Db 661 EAEILFEDIPKEKPPFKRHDVNSTRAQBEVKIPLVNIISLLPKDAQLSLNTLDLQLEH 720
Qy 721 GDITLKGYNLSKALLRSFLNNSQHAKIKNOAIITDETNDLSLVAPOEKQVHKSILPNSLG 780
Db 721 GDITLKGYNLSKALLRSFLNNSQHAKIKNOAIITDETNDLSLVAPOEKQVHKSILPNSLG 780
Qy 781 VSERLQRLTTPPAVSVKVNKGHDQGNPPLDLETTARFRVETHQKTIGGNVTKEKPPSLIV 840
Db 781 VSERLQRLTTPPAVSVKVNKGHDQGNPPLDLETTARFRVETHQKTIGGNVTKEKPPSLIV 840
Qy 841 PLESQMTKEKKTIGKEKENSMEENAEHIGVTEVLLGRKLQHYTDSYLGFLPWEKKKYF 900
Db 841 PLESQMTKEKKTIGKEKENSMEENAEHIGVTEVLLGRKLQHYTDSYLGFLPWEKKKYF 900
Qy 901 QDLLEDEESLKTQLAYFTDSKNTGROLK 928
Db 901 QDLLEDEESLKTQLAYFTDSKNTGROLK 928
```

## RESULT 7

```
US-10-023-894-4
; Sequence 4, Application US/10023894
; Publication No. US20030143669A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: EXPRESSION OF LYSOSOMAL HYDROLASE IN CELLS EXPRESSING PRO-N-
; TITLE OF INVENTION: ACETYLGLUCOSAMINE-1-PHOSPHODIESTER ALPHA-N-ACETYL GLUCOSAMINIDAS
; FILE REFERENCE: 217139US77
; CURRENT APPLICATION NUMBER: US/10/023,894
; CURRENT FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
```

```
; SEQ ID NO 4
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-894-4
```

```
Query Match 100.0%; Score 4907; DB 14; Length 928;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 MLFKLLQRTYTCLSHRYGLYVCFGLGVVVTIIVSAIFQFGEVVLWESRDQHVLFDSYRDMI 60
Db 1 MLFKLLQRTYTCLSHRYGLYVCFGLGVVVTIIVSAIFQFGEVVLWESRDQHVLFDSYRDMI 60
Qy 61 AGKSFQNRCLCLPMPIDVVYTWVNGTDLLELLKELQOVREQMEEEQKAMREILGNKNTTEPTK 120
Db 61 AGKSFQNRCLCLPMPIDVVYTWVNGTDLLELLKELQOVREQMEEEQKAMREILGNKNTTEPTK 120
Qy 121 KSEKQLECLLTHCICKVPMVLDPALPANITLKDVPSLYPSFHSASDIFNVAKPKNPSTNV 180
Db 121 KSEKQLECLLTHCICKVPMVLDPALPANITLKDVPSLYPSFHSASDIFNVAKPKNPSTNV 180
Qy 181 SVVVFSTKDVDEAHSGLLKGNRSQTVWRGVLTTDKVEPGLVLMQDLAFSLGPPPTPKET 240
Db 181 SVVVFSTKDVDEAHSGLLKGNRSQTVWRGVLTTDKVEPGLVLMQDLAFSLGPPPTPKET 240
Qy 241 NQKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKNMTIDGKELTISPA 300
Db 241 NQKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKNMTIDGKELTISPA 300
Qy 301 YLLWDLAISQKODEDISASRFEDNEELRYSLSIERHAPWVRNI FIVTNGQIPSWLNL 360
Db 301 YLLWDLAISQKODEDISASRFEDNEELRYSLSIERHAPWVRNI FIVTNGQIPSWLNL 360
Qy 361 DNPRVTIVTHQDVFRNLHLPTFSSPAIESHHRIEGLSQFIYLNDDVMFGKDVWPDF 420
Db 361 DNPRVTIVTHQDVFRNLHLPTFSSPAIESHHRIEGLSQFIYLNDDVMFGKDVWPDF 420
Qy 421 YSHSGQKQVYLTWPVNCACGPGSWIKDGYCDKACNNSACDWDGDCSGNSGSRVIAG 480
Db 421 YSHSGQKQVYLTWPVNCACGPGSWIKDGYCDKACNNSACDWDGDCSGNSGSRVIAG 480
Qy 481 GGGTSGIGVGHFWPGFQGGGINSVSYCNQGCANSLADKFCDOACNVLSGCFDAGCGGDHF 540
Db 481 GGGTSGIGVGHFWPGFQGGGINSVSYCNQGCANSLADKFCDOACNVLSGCFDAGCGGDHF 540
Qy 541 HELYKVILLPNQTHYIIPKGECLPYFSPAFAVAKRGVEGAYSDNPIIRHASIANKWKTIHL 600
Db 541 HELYKVILLPNQTHYIIPKGECLPYFSPAFAVAKRGVEGAYSDNPIIRHASIANKWKTIHL 600
Qy 601 IMHSGMNAITTHFNLTFTQNTNDEEFKQITVEVDTRGPKLNSTAOQYENLVSPITLLP 660
Db 601 IMHSGMNAITTHFNLTFTQNTNDEEFKQITVEVDTRGPKLNSTAOQYENLVSPITLLP 660
Qy 661 EAEILFEDIPKEKPPFKRHDVNSTRAQBEVKIPLVNIISLLPKDAQLSLNTLDLQLEH 720
Db 661 EAEILFEDIPKEKPPFKRHDVNSTRAQBEVKIPLVNIISLLPKDAQLSLNTLDLQLEH 720
Qy 721 GDITLKGYNLSKALLRSFLNNSQHAKIKNOAIITDETNDLSLVAPOEKQVHKSILPNSLG 780
Db 721 GDITLKGYNLSKALLRSFLNNSQHAKIKNOAIITDETNDLSLVAPOEKQVHKSILPNSLG 780
Qy 781 VSERLQRLTTPPAVSVKVNKGHDQGNPPLDLETTARFRVETHQKTIGGNVTKEKPPSLIV 840
Db 781 VSERLQRLTTPPAVSVKVNKGHDQGNPPLDLETTARFRVETHQKTIGGNVTKEKPPSLIV 840
Qy 841 PLESQMTKEKKTIGKEKENSMEENAEHIGVTEVLLGRKLQHYTDSYLGFLPWEKKKYF 900
Db 841 PLESQMTKEKKTIGKEKENSMEENAEHIGVTEVLLGRKLQHYTDSYLGFLPWEKKKYF 900
Qy 901 QDLLEDEESLKTQLAYFTDSKNTGROLK 928
Db 901 QDLLEDEESLKTQLAYFTDSKNTGROLK 928
```

RESULT 8  
US-10-306-686-1  
; Sequence 1, Application US/10306686  
; Publication No. US20030148460A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM  
; TITLE OF INVENTION: PHOSPHOTESTER ALPHA-GLUCANASE OF THE LYSOSOMAL TARGETING PATHWAY  
; FILE REFERENCE: 230397US77D1V  
; CURRENT APPLICATION NUMBER: US/10/306,686  
; CURRENT FILING DATE: 2002-11-29  
; PRIOR APPLICATION NUMBER: 09/636,596  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-08-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 928  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-306-686-1

Query Match 100.0%; Score 4907; DB 14; Length 928;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFKLQRTYTCLSHRYGLYVCFGLVVTIVSAFQFGEVVLWMSRDQYHVLVDSYRDN 60  
DB 1 MLFKLQRTYTCLSHRYGLYVCFGLVVTIVSAFQFGEVVLWMSRDQYHVLVDSYRDN 60

QY 61 AGKSFQRLCLPMPIDVYVWNGTDLLEKLELQVREOEEQKAMREILGKNTTEPTK 120  
DB 61 AGKSFQRLCLPMPIDVYVWNGTDLLEKLELQVREOEEQKAMREILGKNTTEPTK 120

QY 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSLYPSFHSASDIFNVAKPNPSTNV 180  
DB 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSLYPSFHSASDIFNVAKPNPSTNV 180

QY 181 SVVVFSTKDVEDAHSGLLKGNRSQTVWRGYLTDDKEVGLVMDLAFSLGPPPTKET 240  
DB 181 SVVVFSTKDVEDAHSGLLKGNRSQTVWRGYLTDDKEVGLVMDLAFSLGPPPTKET 240

QY 241 NQKTKLPENLSSKVKLLQLYSEASVALLKNNPKDFQELNKQTKKNTIDGKELTISPA 300  
DB 241 NQKTKLPENLSSKVKLLQLYSEASVALLKNNPKDFQELNKQTKKNTIDGKELTISPA 300

QY 301 YLLWDLAISOSKODEDISASRFEDNEELRYSLRSIERHAPVWRNPIVITNGQIPSWNL 360  
DB 301 YLLWDLAISOSKODEDISASRFEDNEELRYSLRSIERHAPVWRNPIVITNGQIPSWNL 360

QY 361 DNPRTVITVTHQDVFNLSHLPFTSSPAIESHRIEGLSKFYLLNDVWFGKDVWPDF 420  
DB 361 DNPRTVITVTHQDVFNLSHLPFTSSPAIESHRIEGLSKFYLLNDVWFGKDVWPDF 420

QY 421 YSHSGKQVYLTWPVPCAECPGSGWKDGYCDKACNNSACDWDGDCGSCNCGSRYIAG 480  
DB 421 YSHSGKQVYLTWPVPCAECPGSGWKDGYCDKACNNSACDWDGDCGSCNCGSRYIAG 480

QY 481 GGGTSGIVGHPWQFGGGINSVYCNQGCANSWLADKFCDOACNVLSGCGDAGCCQDHF 540  
DB 481 GGGTSGIVGHPWQFGGGINSVYCNQGCANSWLADKFCDOACNVLSGCGDAGCCQDHF 540

QY 541 HELYKVILLNQTIIYIPKGECLPYFSAEVAKRGVEGAYSDNPIIRHASIANKWTIHL 600  
DB 541 HELYKVILLNQTIIYIPKGECLPYFSAEVAKRGVEGAYSDNPIIRHASIANKWTIHL 600

QY 601 IMHSGMATIHFNLITQNTNDEEFKQITVEVDTRGPKLNSTAKQYENLVSPITLLP 660  
DB 601 IMHSGMATIHFNLITQNTNDEEFKQITVEVDTRGPKLNSTAKQYENLVSPITLLP 660

QY 661 EAEILFEDIPEKRPFKRHDVNSTRAQAEVKIPLVNLISLLPKDAQLSLNTLDLQLEH 720

DB 661 EAEILFEDIPEKRPFKRHDVNSTRAQAEVKIPLVNLISLLPKDAQLSLNTLDLQLEH 720  
QY 721 GDITLKGYNLSKALLRSFLMNSQHAKIKNOAITDETNDLSLVAPOEKQVHKSILPNSLG 780  
DB 721 GDITLKGYNLSKALLRSFLMNSQHAKIKNOAITDETNDLSLVAPOEKQVHKSILPNSLG 780  
QY 781 VSERLQRLTTPAVSVKVNHGHDQGNPDLDETTARFRVETHQTKTIGNVTKEPPSLIV 840  
DB 781 VSERLQRLTTPAVSVKVNHGHDQGNPDLDETTARFRVETHQTKTIGNVTKEPPSLIV 840  
QY 841 PLESQMTKEKITGKEKENSMEENAHGIVTEVLLGRKLOHYTDSYGLFPLWEKKKYF 900  
DB 841 PLESQMTKEKITGKEKENSMEENAHGIVTEVLLGRKLOHYTDSYGLFPLWEKKKYF 900  
QY 901 QDLDEESLKTQLAYFTDSKNTGRQLK 928  
DB 901 QDLDEESLKTQLAYFTDSKNTGRQLK 928

## RESULT 9

US-10-023-888-2  
; Sequence 2, Application US/10023888  
; Publication No. US20030119088A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: SOLUBLE GLCNAC PHOSPHOTRANSFERASE  
; FILE REFERENCE: 203515US77  
; CURRENT APPLICATION NUMBER: US/10/023,888  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 1199  
; TYPE: PRT  
; ORGANISM: hybrid  
US-10-023-888-2

Query Match 94.4%; Score 4630; DB 14; Length 1199;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 45 SRDQYHVLVFDYRDNIAKGSFQNLCLPMPIDVYVWNGTDLLELLELQVREOEEQ 104  
DB 35 SRDQYHVLVFDYRDNIAKGSFQNLCLPMPIDVYVWNGTDLLELLELQVREOEEQ 94

QY 105 KAMREILGKNTTEPTKSEKOLECLLTHCIKVPMLVLDPALPANITLKDVPSLYPSFHS 164  
DB 95 KAMREILGKNTTEPTKSEKOLECLLTHCIKVPMLVLDPALPANITLKDVPSLYPSFHS 154

QY 165 SDIFNVAKPNPSTNVSVVVFSTKDVEDAHSGLLKGNRSQTVWRGYLTDDKEVGLVLM 224  
DB 155 SDIFNVAKPNPSTNVSVVVFSTKDVEDAHSGLLKGNRSQTVWRGYLTDDKEVGLVLM 214

QY 225 QDLAFSLGFPPTKETNQLTKLPENLSSKVKLLQLYSEASVALLKNNPKDFQELNKOT 284  
DB 215 QDLAFSLGFPPTKETNQLTKLPENLSSKVKLLQLYSEASVALLKNNPKDFQELNKOT 274

QY 285 KKNMTIDGKELTISPAYLLWDLAISOSKODEDISASRFEDNEELRYSLRSIERHAPVWR 344  
DB 275 KKNMTIDGKELTISPAYLLWDLAISOSKODEDISASRFEDNEELRYSLRSIERHAPVWR 334

QY 345 NIFVITNGQIPSWNLNDNPRVTIYTHQDVFNLSHLPFTSSPAIESHRIEGLSKFYI 404  
DB 335 NIFVITNGQIPSWNLNDNPRVTIYTHQDVFNLSHLPFTSSPAIESHRIEGLSKFYI 394

QY 405 LNDDVWFGKDVWPDFDYSHSGKQVYLTWPVPCAECPGSGWKDGYCDKACNNSACDWD 464  
DB 395 LNDDVWFGKDVWPDFDYSHSGKQVYLTWPVPCAECPGSGWKDGYCDKACNNSACDWD 454

QY 465 GDCSGNSGGSGRYTAGGGTGSIGVGHWPQFGGGINSVYCNQGCANSWLADKFCDOACN 524  
DB 455 GDCSGNSGGSGRYTAGGGTGSIGVGHWPQFGGGINSVYCNQGCANSWLADKFCDOACN 514



525 VLSGFDAGDGGQDHFHLYKVIILLPNQTHYIIPKGECLPYFSPAEVAKRGVEGAYSDNP 584  
Db 515 VLSGFDAGDGGQDHFHLYKVIILLPNQTHYIIPKGECLPYFSPAEVAKRGVEGAYSDNP 574  
Qy 585 IIRHASIANKWKTIHLIMHSGMNAITIHFNLTFTONTDEEFKMQITVEVDTRGPKLNST 644  
Db 575 IIRHASIANKWKTIHLIMHSGMNAITIHFNLTFTONTDEEFKMQITVEVDTRGPKLNST 634  
Qy 645 AOKGYENLVSPITLLPEAEILFEDIPKEKRPFKRHDVNSTRAQEEVKIPLVNIISLLP 704  
Db 635 AOKGYENLVSPITLLPEAEILFEDIPKEKRPFKRHDVNSTRAQEEVKIPLVNIISLLP 694  
Qy 705 KDAQLSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQAKIKNOAIITDETNDLSVA 764  
Db 695 KDAQLSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQAKIKNOAIITDETNDLSVA 754  
Qy 765 POEKQVHKSILPNSLGVSERLQRLTFPAVSVKVNGHDQGNPPLDLETTARFRVETHTQK 824  
Db 755 POEKQVHKSILPNSLGVSERLQRLTFPAVSVKVNGHDQGNPPLDLETTARFRVETHTQK 814  
Qy 825 TIGGNVTEKPPSLIVPLESQMTKEKKTGKEKNSRMEENAHNIGVTEVLLGRKLOHY 884  
Db 815 TIGGNVTEKPPSLIVPLESQMTKEKKTGKEKNSRMEENAHNIGVTEVLLGRKLOHY 874  
Qy 885 TDSYLGFLPWKKYFQDLDDEESLKTQLAYFTDSKNTR 925  
Db 875 TDSYLGFLPWKKYFQDLDDEESLKTQLAYFTDSKNTR 915

RESULT 10  
US-10-023-889-2  
; Sequence 2, Application US/10023889  
; Publication No. US20030124652A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS OF PRODUCING HIGH MANNOSE GLYCOPROTEINS IN COMPLEX CARBOH  
; TITLE OF INVENTION: DEFICIENT CELLS  
; FILE REFERENCE: 2035120S77  
; CURRENT APPLICATION NUMBER: US/10/023.889  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 1199  
; TYPE: PRT  
; ORGANISM: hybrid  
US-10-023-889-2

Query Match 94.4%; Score 4630; DB 14; Length 1199;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 45 SRDQYHVLFDSDYRDNIAKSGFQNRCLPMPIDVYVYVWNGTDLLELKLQOVRQMBEEQ 104  
Db 35 SRDQYHVLFDSDYRDNIAKSGFQNRCLPMPIDVYVYVWNGTDLLELKLQOVRQMBEEQ 94  
Qy 105 KAMREILGKNTTEPTKKSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSLYPSFSHA 164  
Db 95 KAMREILGKNTTEPTKKSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSLYPSFSHA 154  
Qy 165 SDIFNVAKPKNPSTNVSVVVFDSKDVDAHSGLLKGNRSQTVWRGYLTDTKEVPGVLVM 224  
Db 155 SDIFNVAKPKNPSTNVSVVVFDSKDVDAHSGLLKGNRSQTVWRGYLTDTKEVPGVLVM 214  
Qy 225 QDLAPLSGFPPTFKETNQLTKLPENLSKVKLLQLYSEASVALLKNNPKDFOLNKT 284  
Db 215 QDLAPLSGFPPTFKETNQLTKLPENLSKVKLLQLYSEASVALLKNNPKDFOLNKT 274  
Qy 285 KKNMTIDGKELTISPAYLLWDLISQSKQDEDISASRFEDNEELRYSLRSIERHAPWR 344  
Db 275 KKNMTIDGKELTISPAYLLWDLISQSKQDEDISASRFEDNEELRYSLRSIERHAPWR 334

345 NPIFVTNGOIPSWNLNDNPRVTIVTHQDVFRNLSHLPFPSSPAIESHIHRIEGLSOKFY 404  
Db 335 NPIFVTNGOIPSWNLNDNPRVTIVTHQDVFRNLSHLPFPSSPAIESHIHRIEGLSOKFY 394  
Qy 405 LNDVNMFGKVDWPDYSHSGKQKYYLTWPVPCNCAEGCPGSGWIKDGYCDKACNNSACDWD 464  
Db 395 LNDVNMFGKVDWPDYSHSGKQKYYLTWPVPCNCAEGCPGSGWIKDGYCDKACNNSACDWD 454  
Qy 465 GDCSCNSGSGRYIAGGGGTGSIIGVCHPWQFGGGSINVSVCNQGCSANSLADKFCDOACN 524  
Db 455 GDCSCNSGSGRYIAGGGGTGSIIGVCHPWQFGGGSINVSVCNQGCSANSLADKFCDOACN 514  
Qy 525 VLSGFDAGDGGQDHFHLYKVIILLPNQTHYIIPKGECLPYFSPAEVAKRGVEGAYSDNP 584  
Db 515 VLSGFDAGDGGQDHFHLYKVIILLPNQTHYIIPKGECLPYFSPAEVAKRGVEGAYSDNP 574  
Qy 585 IIRHASIANKWKTIHLIMHSGMNAITIHFNLTFTONTDEEFKMQITVEVDTRGPKLNST 644  
Db 575 IIRHASIANKWKTIHLIMHSGMNAITIHFNLTFTONTDEEFKMQITVEVDTRGPKLNST 634  
Qy 645 AOKGYENLVSPITLLPEAEILFEDIPKEKRPFKRHDVNSTRAQEEVKIPLVNIISLLP 704  
Db 635 AOKGYENLVSPITLLPEAEILFEDIPKEKRPFKRHDVNSTRAQEEVKIPLVNIISLLP 694  
Qy 705 KDAQLSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQAKIKNOAIITDETNDLSVA 764  
Db 695 KDAQLSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQAKIKNOAIITDETNDLSVA 754  
Qy 765 POEKQVHKSILPNSLGVSERLQRLTFPAVSVKVNGHDQGNPPLDLETTARFRVETHTQK 824  
Db 755 POEKQVHKSILPNSLGVSERLQRLTFPAVSVKVNGHDQGNPPLDLETTARFRVETHTQK 814  
Qy 825 TIGGNVTEKPPSLIVPLESQMTKEKKTGKEKNSRMEENAHNIGVTEVLLGRKLOHY 884  
Db 815 TIGGNVTEKPPSLIVPLESQMTKEKKTGKEKNSRMEENAHNIGVTEVLLGRKLOHY 874  
Qy 885 TDSYLGFLPWKKYFQDLDDEESLKTQLAYFTDSKNTR 925  
Db 875 TDSYLGFLPWKKYFQDLDDEESLKTQLAYFTDSKNTR 915

RESULT 11  
US-10-023-890-2  
; Sequence 2, Application US/10023890  
; Publication No. US20030124653A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHOD OF PRODUCING GLYCOPROTEINS HAVING REDUCED COMPLEX CARBOH  
; TITLE OF INVENTION: MAMMALIAN CELLS  
; FILE REFERENCE: 203510US77  
; CURRENT APPLICATION NUMBER: US/10/023.890  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 1199  
; TYPE: PRT  
; ORGANISM: hybrid  
US-10-023-890-2

Query Match 94.4%; Score 4630; DB 14; Length 1199;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 45 SRDQYHVLFDSDYRDNIAKSGFQNRCLPMPIDVYVYVWNGTDLLELKLQOVRQMBEEQ 104  
Db 35 SRDQYHVLFDSDYRDNIAKSGFQNRCLPMPIDVYVYVWNGTDLLELKLQOVRQMBEEQ 94  
Qy 105 KAMREILGKNTTEPTKKSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSLYPSFSHA 164  
Db 95 KAMREILGKNTTEPTKKSEKQLECLLTHCIKVPMLVLDPALPANITLKDVPSLYPSFSHA 154  
Qy 165 SDIFNVAKPKNPSTNVSVVVFDSKDVDAHSGLLKGNRSQTVWRGYLTDTKEVPGVLVM 224

Db	155	SDIFNVAKPKNPSTNVS	VVFDSTK	VEDAHSGLLKGN	SRQTVMRG	YLTITDK	VPGLVM	214																		
Qy	225	QDLAFLSGFPPTFKET	NQLTKL	PENLSSKV	KLQLYSEAS	VALLKLN	PKPQELNKOT	284																		
Db	215	QDLAFLSGFPPTFKET	NQLTKL	PENLSSKV	KLQLYSEAS	VALLKLN	PKPQELNKOT	274																		
Qy	285	KKNMTIDGKELTISP	AYLLWDL	SAISQKODE	BISASR	FEDNEEL	RYSLRSIRHAPWR	344																		
Db	275	KKNMTIDGKELTISP	AYLLWDL	SAISQKODE	BISASR	FEDNEEL	RYSLRSIRHAPWR	334																		
Qy	345	NIFIVTNGQIPSW	NLNDNPRVT	I	THQDV	FRNL	SHLPTFSSPAIESH	IRIEGLSQKFIY	404																	
Db	335	NIFIVTNGQIPSW	NLNDNPRVT	I	THQDV	FRNL	SHLPTFSSPAIESH	VRHIEGLSQKFIY	394																	
Qy	405	LNDVFMFGKDW	PPDDPYSHS	KGKVYLT	W	PVNC	ASGCPGSGWKDGYCD	KACNNSACDWD	464																	
Db	395	LNDVFMFGKDW	PPDDPYSHS	KGKVYLT	W	PVNC	ASGCPGSGWKDGYCD	KACNNSACDWD	454																	
Qy	465	GGDCSGSGSRY	IAGGGTSG	IGVGH	PWQFGG	I	NSVSYCNOGCANS	WLADKFCDOACN	524																	
Db	455	GGDCSGSGSRY	IAGGGTSG	IGVGH	PWQFGG	I	NSVSYCNOGCANS	WLADKFCDOACN	514																	
Qy	525	VLSCGFADGCGD	HFEHLYKV	ILLPNQ	THYIIP	KGECLPY	FSFAE	VAKRGV	GAYS	DNP	584															
Db	515	VLSCGFADGCGD	HFEHLYKV	ILLPNQ	THYIIP	KGECLPY	FSFAE	VAKRGV	GAYS	DNP	574															
Qy	585	IIRHASIANKWK	THILM	HSGMNATT	HNLT	FNQNTN	DEFKMOIT	VEVD	TREG	PKLNST	644															
Db	575	IIRHASIANKWK	THILM	HSGMNATT	HNLT	FNQNTN	DEFKMOIT	VEVD	TREG	PKLNST	634															
Qy	645	AQKGVENLVS	PITLLPEAE	ILFEDIP	KEKRF	PKFKR	HVNSTR	RAOE	EVKIP	LNVNISLLP	704															
Db	635	AQKGVENLVS	PITLLPEAE	ILFEDIP	KEKRF	PKFKR	HVNSTR	RAOE	EVKIP	LNVNISLLP	694															
Qy	705	KDAQLSANT	LDLQLEH	DTLKG	YNLSK	SALLBS	FLMNSO	HAKIKNO	QAIIT	DETNDLSVA	764															
Db	695	KDAQLSANT	LDLQLEH	DTLKG	YNLSK	SALLBS	FLMNSO	HAKIKNO	QAIIT	DETNDLSVA	754															
Qy	765	PQEKQVHKS	ILPNS	LGVSR	ERLQ	LT	FP	AVS	VKN	GHDQGN	824															
Db	755	PQEKQVHKS	ILPNS	LGVSR	ERLQ	LT	FP	AVS	VKN	GHDQGN	814															
Qy	825	TIGGNVTK	PKPSSLIV	PL	ESQMT	KEKKIT	G	KE	N	S	RM	EEN	AEN	HIG	V	TE	V	L	G	R	K	L	O	H	Y	884
Db	815	TIGGNVTK	PKPSSLIV	PL	ESQMT	KEKKIT	G	KE	N	S	RM	EEN	AEN	HIG	V	TE	V	L	G	R	K	L	O	H	Y	874
Qy	885	TDSYLG	FLPWEK	KKYPQD	LLD	DEE	S	L	K	T	O	L	A	Y	F	T	D	S	K	N	T	G	R	925		
Db	875	TDSYLG	FLPWEK	KKYPQD	LLD	DEE	S	L	K	T	O	L	A	Y	F	T	D	S	K	N	T	G	R	915		

RESULT 12  
US-10-024-197-2  
; Sequence 2, Application US/10024197  
; Publication No. US2003013924A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: HIGHLY PHOSPHORYLATED ACID BETA-GLUCOCEREBROSIDASE AND METHODS  
; TITLE OF INVENTION: TREATING GAUCHER'S DISEASE  
; FILE REFERENCE: 209794US0

## RESULT 13

US-10-023-894-2

; Sequence 2, Application US/10023894

; Publication No. US20030143669A1

; GENERAL INFORMATION:

; APPLICANT: CANFIELD, William

; APPLICANT: KORNFIELD, Stuart

; TITLE OF INVENTION: EXPRESSION OF LYOSOMAL HYDROLASE IN CELLS EXPRESSING PRO-N-

; TITLE OF INVENTION: ACETYLGUCOSAMINE-1-PHOSPHODIESTER ALPHA-N-ACETYL GLUCOSAMINIDAS

; FILE REFERENCE: 217139US77

## RESULT 12

US-10-024-197-2

; Sequence 2, Application US/10024197

; Publication No. US2003013924A1

; GENERAL INFORMATION:

; APPLICANT: CANFIELD, William

; APPLICANT: KORNFIELD, Stuart

; TITLE OF INVENTION: HIGHLY PHOSPHORYLATED ACID BETA-GLUCOCEREBROSIDASE AND METHODS

; TITLE OF INVENTION: TREATING GAUCHER'S DISEASE

; FILE REFERENCE: 209794US0

; CURRENT APPLICATION NUMBER: US/10/024,197

; CURRENT FILING DATE: 2001-12-21

; NUMBER OF SEQ ID NOS: 27

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 2

; LENGTH: 1199

; TYPE: PRT

; ORGANISM: hybrid

US-10-024-197-2

Query Match 94.4%; Score 4630; DB 14; Length 1199;

Best Local Similarity 99.3%; Pred. No. 0;

; CURRENT APPLICATION NUMBER: US/10/023,894  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: Patent version 3.1  
; SEQ ID NO 2  
; LENGTH: 1199  
; TYPE: PRT  
; ORGANISM: hybrid  
US-10-023-894-2

Query Match 94.4%; Score 4630; DB 14; Length 1199;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 45 SRDQVHVLFDSDYRDNIAKSCFNRLCLPMPIDVVTWNGTDLLELLKELQVREOMEEQ 104  
Db 35 SRDQVHVLFDSDYRDNIAKSCFNRLCLPMPIDVVTWNGTDLLELLKELQVREOMEEQ 94  
QY 105 KAMREILGKNTTEPTKSEKOLECLTHCIKVPMLVLDPALPANITLKDPSLPSPHSA 164  
Db 95 KAMREILGKNTTEPTKSEKOLECLTHCIKVPMLVLDPALPANITLKDPSLPSPHSA 154  
QY 165 SDIFNVAKPKNPSTNVSVVVDSTKQVEDAHSGLLKGNRSQTVWRGYLTDDKEVPGVLVM 224  
Db 155 SDIFNVAKPKNPSTNVSVVVDSTKQVEDAHSGLLKGNRSQTVWRGYLTDDKEVPGVLVM 214  
QY 225 QDLAFLSGFPPTFKETNOLTKLPENLSSKVKLLQLYSEASVALLKNNPKDFOLNKOT 284  
Db 215 QDLAFLSGFPPTFKETNOLTKLPENLSSKVKLLQLYSEASVALLKNNPKDFOLNKOT 274  
QY 285 KKNMTIDGKELTISPAYLLDLSAISQSKQEDISASRFEDNEELRYSLRSIERHAPVVR 344  
Db 275 KKNMTIDGKELTISPAYLLDLSAISQSKQEDISASRFEDNEELRYSLRSIERHAPVVR 334  
QY 345 NIFVTNGQIPSWNLNDNPRVTIYTHQDVFNLSHLPTSSPATESHIHIEGLSKFIY 404  
Db 335 NIFVTNGQIPSWNLNDNPRVTIYTHQDVFNLSHLPTSSPATESHIHIEGLSKFIY 394  
QY 405 LNDVVRGKQVDPDFYSHSGKQVYLTWPVPCNCAEGCPGQSWIKDGYCDXACNNSACDWD 464  
Db 395 LNDVVRGKQVDPDFYSHSGKQVYLTWPVPCNCAEGCPGQSWIKDGYCDXACNNSACDWD 454  
QY 465 GDCSGNSGGRYTAGGGTSGISGVHPWQGGINSVSCNQCANSWLADKFCDOACN 524  
Db 455 GDCSGNSGGRYTAGGGTSGISGVHPWQGGINSVSCNQCANSWLADKFCDOACN 514  
QY 525 VLSGCFDAGDQGHFHELYKIVILLPQTHVIIIPKGECLPYFSPAFAVAKRGVEGAYSDNP 584  
Db 515 VLSGCFDAGDQGHFHELYKIVILLPQTHVIIIPKGECLPYFSPAFAVAKRGVEGAYSDNP 574  
QY 585 IIRHASIANKWKTTHLIMHSGMNAITTHFNLTQNTNDEBFKQITVEVDTRGPKLNT 644  
Db 575 IIRHASIANKWKTTHLIMHSGMNAITTHFNLTQNTNDEBFKQITVEVDTRGPKLNT 634  
QY 645 AOKGVENLVSPITLLPAEILFEDIPEKRPFPKFRHDVNSTRAQAEVKIPLVNTSLLP 704  
Db 635 AOKGVENLVSPITLLPAEILFEDIPEKRPFPKFRHDVNSTRAQAEVKIPLVNTSLLP 694  
QY 705 KDAQLSNTLDLQLEHGDITLKGYNLSKALLRSFLANSQHAKIKNOAIIITDETNDLSVA 764  
Db 695 KDAQLSNTLDLQLEHGDITLKGYNLSKALLRSFLANSQHAKIKNOAIIITDETNDLSVA 754  
QY 765 POEKQVHKSILPNSLGVSERLQRTFFAVSVKNGHDOGQNPPLDLETTARFVETHQK 824  
Db 755 POEKQVHKSILPNSLGVSERLQRTFFAVSVKNGHDOGQNPPLDLETTARFVETHQK 814  
QY 825 TIGGNVTKEKPPSLIIVPLESOMTKEKKTGKEKNSRMEENHIGVTEVLLGRKLOHY 884  
Db 815 TIGGNVTKEKPPSLIIVPLESOMTKEKKTGKEKNSRMEENHIGVTEVLLGRKLOHY 874  
QY 885 TDSYLGFLPWEKKYFQDLDEBSLTKQLAYFTDSKNTGR 925  
Db 875 TDSYLGFLPWEKKYFQDLDEBSLTKQLAYFTDSKNTGR 915

RESULT 14  
US-10-120-801-88  
; Sequence 88, Application US/10120801  
; Publication No. US20030203843A1  
; GENERAL INFORMATION:  
; APPLICANT: Pena, Carol  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Padigar, Muralidhara  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Spytak, Kimberly  
; APPLICANT: Mehraban, Fuad  
; APPLICANT: Topper, James N.  
; APPLICANT: Malyankar, Uriel  
; APPLICANT: Wasserman, Scott  
; APPLICANT: Edinger, Salomit  
; APPLICANT: Smithson, Glennda  
; APPLICANT: Gunther, Erik  
; APPLICANT: Komuves, Laszlo  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-340  
; CURRENT APPLICATION NUMBER: US/10/120,801  
; CURRENT FILING DATE: 2002-04-11  
; PRIOR APPLICATION NUMBER: 60/285748  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: 60/286068  
; PRIOR FILING DATE: 2001-04-24  
; PRIOR APPLICATION NUMBER: 60/286292  
; PRIOR FILING DATE: 2001-04-25  
; PRIOR APPLICATION NUMBER: 60/288334  
; PRIOR FILING DATE: 2001-05-03  
; PRIOR APPLICATION NUMBER: 60/291241  
; PRIOR FILING DATE: 2001-05-16  
; PRIOR APPLICATION NUMBER: 60/322284  
; PRIOR FILING DATE: 2001-09-14  
; PRIOR APPLICATION NUMBER: 60/285609  
; PRIOR FILING DATE: 2001-04-20  
; NUMBER OF SEQ ID NOS: 155  
; SOFTWARE: Patent Ver. 2.1  
; SEQ ID NO 88  
; LENGTH: 847  
; TYPE: PRT  
; ORGANISM: human  
US-10-120-801-88

Query Match 91.2%; Score 4473; DB 14; Length 847;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 845; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLFKLQRTYTCLSHRYGLYVCFGLVVVTIVSAFQFGEVLEWSDQYHVLFDSDYRDN 60  
Db 1 MLFKLQRTYTCLSHRYGLYVCFGLVVVTIVSAFQFGEVLEWSDQYHVLFDSDYRDN 60  
QY 61 AGKSTQNRCLPMPIDVVTWNGTDLLELLKELQVREOMEEQKAMREILGKNTTEPTK 120  
Db 61 AGKSTQNRCLPMPIDVVTWNGTDLLELLKELQVREOMEEQKAMREILGKNTTEPTK 120  
QY 121 KSEKOLECLLTHCIKVPMLVLDPALPANITLKDPSLPSPHSAASDIFNVAKPKNPSTNV 180  
Db 121 KSEKOLECLLTHCIKVPMLVLDPALPANITLKDPSLPSPHSAASDIFNVAKPKNPSTNV 180  
QY 181 SVVVFDTSTKQVEDAHSGLLKGNRSQTVWRGYLTDDKEVPGVLVMQDLAFLSGFPPTFKET 240  
Db 181 SVVVFDTSTKQVEDAHSGLLKGNRSQTVWRGYLTDDKEVPGVLVMQDLAFLSGFPPTFKET 240  
QY 241 NQLKTKLPENLSSKVKLLQLYSEASVALLKNNPKDFOLNKQTKKNTIDCKELTISPA 300  
Db 241 NQLKTKLPENLSSKVKLLQLYSEASVALLKNNPKDFOLNKQTKKNTIDCKELTISPA 300  
QY 301 YLLWDLISAISQSKQEDISASRFEDNEELRYSLRSIERHAPVVRNIFVTNGQIPSWNLN 360

Db 301 YLLWDLSSAISQSKODEDISASRFEDNEELRYLSRISIERHAPVVRNFIIVTNGQIPSWLNL 360  
Qy 361 DNPRTVITVHQDVFRNLNLSHPTSSPAIESHIIHRIEGLSQKFIYLLNDDVNFQKDVWPDFF 420  
Db 361 DNPRTVITVHQDVFRNLNLSHPTSSPAIESHIIHRIEGLSQKFIYLLNDDVNFQKDVWPDFF 420  
Qy 421 YSHSGQKQVLTWPVPCACGCGSWIKDGYCDKACNNSACDWDGDCGNSGGSRYIAG 480  
Db 421 YSHSGQKQVLTWPVPCACGCGSWIKDGYCDKACNNSACDWDGDCGNSGGSRYIAG 480  
Qy 481 GGGTSGSIVGQHPWQFGGINSVSCNOCGANSWLADKFCDOACNVLSCGFDAGDCQDHF 540  
Db 481 GGGTSGSIVGQHPWQFGGINSVSCNOCGANSWLADKFCDOACNVLSCGFDAGDCQDHF 540  
Qy 541 HELYKVILLPNQTHYIIPKGECLPYFSPFAEVAKEGVSAGSDNPIIRHASIANKWKTIHL 600  
Db 541 HELYKVILLPNQTHYIIPKGECLPYFSPFAEVAKEGVSAGSDNPIIRHASIANKWKTIHL 600  
Qy 601 IMHSGMNATTHFNLTQNTNDEEFKQIITVEVDTRREGPKLNSTAKQGYENLVSPITLLP 660  
Db 601 IMHSGMNATTHFNLTQNTNDEEFKQIITVEVDTRREGPKLNSTAKQGYENLVSPITLLP 660  
Qy 661 EAEILFEDIPEKRPFPKRDHVNSTRRAQEEVKIPLVNIISLLPKDAQSLNLTDLQLEH 720  
Db 661 EAEILFEDIPEKRPFPKRDHVNSTRRAQEEVKIPLVNIISLLPKDAQSLNLTDLQLEH 720  
Qy 721 GDIITLKGYNLSKALLRSFLMNSOHAKIKNOAIITDETNDLSLVAPOEKQVHKSILPNSLG 780  
Db 721 GDIITLKGYNLSKALLRSFLMNSOHAKIKNOAIITDETNDLSLVAPOEKQVHKSILPNSLG 780  
Qy 781 VSELRQRLTTPAVSVKNGHDQGNPPDLLETTARFRVETHTOKTIGGNVTKKPPSLIV 840  
Db 781 VSELRQRLTTPAVSVKNGHDQGNPPDLLETTARFRVETHTOKTIGGNVTKKPPSLIV 840  
Qy 841 PLESQM 846  
Db 841 PLESQM 846

RESULT 15  
US-10-094-749-2392  
; Sequence 2392, Application US/10094749  
; Publication No. US20030219741A1  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI  
; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: YAMAMOTO, JUN-ICHI  
; APPLICANT: ISONO, YUUKO  
; APPLICANT: HIO, YURI  
; APPLICANT: OTSUKA, KAORU  
; APPLICANT: NAGAI, KETICHI  
; APPLICANT: IRIE, RYOTARO  
; APPLICANT: TAMECHIKA, ICHIRO  
; APPLICANT: SEKI, NAOHICO  
; APPLICANT: YOSHIKAWA, TSUTOMU  
; APPLICANT: OTSUKA, MOTOKYU  
; APPLICANT: NAGAHARI, KENJI  
; APPLICANT: MASUHO, YASUHIKO  
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA  
; FILE REFERENCE: 084335/0160  
; CURRENT APPLICATION NUMBER: US/10/094,749  
; CURRENT FILING DATE: 2002-03-12  
; PRIOR APPLICATION NUMBER: 60/350,435  
; PRIOR FILING DATE: 2002-01-24  
; PRIOR APPLICATION NUMBER: JP 2001-328381  
; PRIOR FILING DATE: 2001-09-14  
; NUMBER OF SEQ ID NOS: 3381  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 2392

; LENGTH: 846  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-094-749-2392  
Query Match 91.1%; Score 4468; DB 14; Length 846;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 844; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 MLFKLLQROTTCLSHRYGLVCFGLGVVVTTSVSAFQGEVVLWSRDQYHVLFDSDYRDN 60  
Db 1 MLFKLLQROTTCLSHRYGLVCFGLGVVVTTSVSAFQGEVVLWSRDQYHVLFDSDYRDN 60  
Qy 61 AGKSFQNRCLCPMPIDVVVYTWVNGTDLLELLKELQVREQMEEEKAMREILKNTTTEPTK 120  
Db 61 AGKSFQNRCLCPMPIDVVVYTWVNGTDLLELLKELQVREQMEEEKAMREILKNTTTEPTK 120  
Qy 121 KSEKQLECLLTHCIKVPMLVLDPALPANITIKDVPISYPSPHSASDIFNVAKPKNPSTNV 180  
Db 121 KSEKQLECLLTHCIKVPMLVLDPALPANITIKDVPISYPSPHSASDIFNVAKPKNPSTNV 180  
Qy 181 SVVVFDSKQVEDAHSGLLKGNRSQTVMRGYLTDDKEVPGLVLMQDLAFISGPPPTFKET 240  
Db 181 SVVVFDSKQVEDAHSGLLKGNRSQTVMRGYLTDDKEVPGLVLMQDLAFISGPPPTFKET 240  
Qy 241 NQKTKLPENLSSKVKLLQLLYSEASVALLKLNPKDFQELNKKTKKNTIDGKELTISPA 300  
Db 241 NQKTKLPENLSSKVKLLQLLYSEASVALLKLNPKDFQELNKKTKKNTIDGKELTISPA 300  
Qy 301 YLLWDLSSAISQSKODEDISASRFEDNEELRYLSRISIERHAPVVRNFIIVTNGQIPSWLNL 360  
Db 301 YLLWDLSSAISQSKODEDISASRFEDNEELRYLSRISIERHAPVVRNFIIVTNGQIPSWLNL 360  
Qy 361 DNPRTVITVHQDVFRNLNLSHPTSSPAIESHIIHRIEGLSQKFIYLLNDDVNFQKDVWPDFF 420  
Db 361 DNPRTVITVHQDVFRNLNLSHPTSSPAIESHIIHRIEGLSQKFIYLLNDDVNFQKDVWPDFF 420  
Qy 421 YSHSGQKQVLTWPVPCACGCGSWIKDGYCDKACNNSACDWDGDCGNSGGSRYIAG 480  
Db 421 YSHSGQKQVLTWPVPCACGCGSWIKDGYCDKACNNSACDWDGDCGNSGGSRYIAG 480  
Qy 481 GGGTSGSIVGQHPWQFGGINSVSCNOCGANSWLADKFCDOACNVLSCGFDAGDCQDHF 540  
Db 481 GGGTSGSIVGQHPWQFGGINSVSCNOCGANSWLADKFCDOACNVLSCGFDAGDCQDHF 540  
Qy 541 HELYKVILLPNQTHYIIPKGECLPYFSPFAEVAKEGVSAGSDNPIIRHASIANKWKTIHL 600  
Db 541 HELYKVILLPNQTHYIIPKGECLPYFSPFAEVAKEGVSAGSDNPIIRHASIANKWKTIHL 600  
Qy 601 IMHSGMNATTHFNLTQNTNDEEFKQIITVEVDTRREGPKLNSTAKQGYENLVSPITLLP 660  
Db 601 IMHSGMNATTHFNLTQNTNDEEFKQIITVEVDTRREGPKLNSTAKQGYENLVSPITLLP 660  
Qy 661 EAEILFEDIPEKRPFPKRDHVNSTRRAQEEVKIPLVNIISLLPKDAQSLNLTDLQLEH 720  
Db 661 EAEILFEDIPEKRPFPKRDHVNSTRRAQEEVKIPLVNIISLLPKDAQSLNLTDLQLEH 720  
Qy 721 GDIITLKGYNLSKALLRSFLMNSOHAKIKNOAIITDETNDLSLVAPOEKQVHKSILPNSLG 780  
Db 721 GDIITLKGYNLSKALLRSFLMNSOHAKIKNOAIITDETNDLSLVAPOEKQVHKSILPNSLG 780  
Qy 781 VSELRQRLTTPAVSVKNGHDQGNPPDLLETTARFRVETHTOKTIGGNVTKKPPSLIV 840  
Db 781 VSELRQRLTTPAVSVKNGHDQGNPPDLLETTARFRVETHTOKTIGGNVTKKPPSLIV 840  
Qy 841 PLESQM 846  
Db 841 PLESQM 846

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2004, 12:39:03 ; Search time 28.3503 Seconds  
(without alignments)  
3149.495 Million cell updates/sec

Title: US-10-023-888-4  
Perfect score: 4907  
Sequence: 1 MLFKLQRTVTCLSHRYGL.....SLKTLQLAYFTDSKNTGRQLK 928

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79:.\*  
1: Pir1.\*  
2: Pir2.\*  
3: Pir3.\*  
4: Pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	734.5	15.0	384	2 I49528	hypothetical prote
2	322.5	6.6	586	2 T34703	hypothetical prote
3	314	6.4	541	2 T34701	hypothetical prote
4	266	5.4	545	2 C82014	probable capsule b
5	256	5.2	532	2 G70536	probable cpsy prot
6	255.5	5.2	542	2 D87185	probable UDP-gluc
7	228.5	4.7	373	2 S42434	hypothetical prote
8	175.5	3.6	2352	2 T30201	Notch homolog prot
9	166.5	3.4	2471	2 A49128	cell-fate determin
10	166.5	3.4	1429	2 S06434	homeotic protein 1
11	165	3.4	2531	2 S18188	Notch protein homo
12	163	3.4	2531	2 A46019	Notch-1 protein -
13	162.5	3.3	2437	2 S42612	transmembrane prot
14	154.5	3.1	2321	2 T85549	Notch3 protein - h
15	152	3.1	1964	2 T05059	Notch4 - mouse
16	148	3.0	861	2 A48825	Notch homolog Motc
17	147	3.0	1203	2 A49175	Notch B protein -
18	147	3.0	2318	2 S45306	Notch 3 protein -
19	146.5	3.0	387	2 A49175	Notch A protein -
20	146	3.0	2531	2 T31070	Notch homolog - se
21	145.5	3.0	2555	2 A40043	Notch protein homo
22	144.5	2.9	1755	2 S70298	TYB protein - years
23	140.5	2.9	1755	2 S50663	TYB protein - years
24	138	2.8	2524	2 A35844	Notch protein - Af
25	137.5	2.8	1755	2 S65986	TYB protein - years
26	136.5	2.8	1701	2 A26868	major merozoite su
27	135.5	2.8	1701	2 A54498	major merozoite su
28	134	2.7	1295	2 A32901	gip1 protein precu
29	133.5	2.7	540	2 T34702	Hypothetical prote

30	133.5	2.7	1969	2 T08875	histidine kinase h
31	133.5	2.7	2052	2 C97038	phage-related prot
32	131.5	2.7	957	2 C69463	type I restriction
33	131.5	2.7	1422	2 T18404	chromatin remodel
34	131.5	2.7	1640	2 A24594	probable major sur
35	130	2.6	484	2 T16595	hypothetical prote
36	129.5	2.6	1639	2 S05603	major merozoite su
37	129	2.6	1276	2 T09204	probable tail-host
38	129	2.6	2703	1 A24420	Notch protein - fr
39	128.5	2.6	756	2 C64236	hypothetical prote
40	128	2.6	644	2 T47835	DNA mismatch repai
41	127.5	2.6	780	2 B70112	TYB protein - years
42	127.5	2.6	1755	2 S69963	hypothetical prote
43	127	2.6	1459	2 S50437	TYB protein - years
44	126.5	2.6	1328	2 S52894	protein kinesin p2
45	125.5	2.6	987	2 B96766	

ALIGNMENTS

RESULT 1  
I49528  
hypothetical protein - mouse (fragment)  
N;Alternate names: CDNA1  
C;Species: Mus musculus (house mouse)  
C;Date: 09-Mar-1996 #sequence\_revision 09-Mar-1996 #text\_change 09-Jul-2004  
C;Accession: I49528  
R;Cordes, S.P.; Barsh, G.S.  
Cell 79, 1025-1034, 1994  
A;Title: The mouse segmentation gene kr encodes a novel basic domain-leucine zipper tr  
A;Reference number: A55200; MUID:95094266; PMID:8001130  
A;Accession: I49528  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-384 <RES>  
A;Cross-references: UNIPROT:Q61340; GB:IJ36434; NID:9625041; PIDN:AAA65688.1; PID:96250

Query Match 15.0%; Score 734.5; DB 2; Length 384;  
Best Local Similarity 57.9%; Pred. No. 2.1e-38;  
Matches 165; Conservative 28; Mismatches 71; Indels 21; Gaps 6;

Qy	644	TAAQGVENLAVSPITLLPERAILPEDIPEKRPKFRHDVNSTRAQEVKIPLVNISLL	703
Db	3	TTQKAYESLVSVPTLPQADVFEDVPKEKRPKIRHDVNAIGRQEVKIPRVNISLL	62
Qy	704	PKDAQSLNLTLDLQLEHGDITLKGYNLSKALLRSFLMNSQHAKIKNOAIITDETNDLSLV	763
Db	63	PXEAQVRLSNLDLQLEHGDITLKGYNLSKALLRSFLGNSLDTKIKPCA-RDETCKGNLE	121
Qy	764	APQEQVHKSIILPNSLGVSRQLRITFFPAVSVKVNGHDGQGNPPDLDTTARFRVETHQ	823
Db	122	VPQENPSHRR--PHGFAGSHRSERWTAFATVTVKGRDHALNPPFPVLETNARL----	174
Qy	824	KTIGGNVTKEKPPSLIVPLESQMTKEKITGKEKNSRMEENAHNIGVTEVLGRKLQH	883
Db	175	PILGTVTSKENSLPIVPPESHLP-----KEESDRAEGNA---VPVKELVPORCSR	224
Qy	884	YTDSYLGFLPWKKYFQDLDERESLKTQLAYFTDSKNTGRQLK	928
Db	225	IIQA---FCPGKKKKYFQDLDAEESLKTQLAYFTDSKNTGRQLK	266

RESULT 2  
T34703  
hypothetical protein SC1C3.11 SC1C3.11 - Streptomyces coelicolor  
C;Species: Streptomyces coelicolor  
C;Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 09-Jul-2004  
C;Accession: T34703  
R;Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, May 1998  
A;Reference number: Z21554  
A;Accession: T34703





RESULT 6  
D87185  
Probable UDP-glucose-4-epimerase [imported] - Mycobacterium leprae  
C:Species: Mycobacterium leprae  
C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004  
C:Accession: D87185  
R: Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; H  
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,  
eam, M.A.; Rutherford, K.M.  
Nature 409, 1007-1011, 2001  
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq  
A:Title: Massive gene decay in the leprosy bacillus.  
A:Reference number: A86909; MUID:21128732; PMID:11234002  
A:Accession: D87185  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-542 <STO>  
A:Cross-references: UNIPROT:Q50025; GB:AL450380; NID:gl3093740; PIDN:CAC31164.1; GSPDB:G  
C:Genetics:  
A:Gene: cpsY

Query Match 5.2%; Score 255.5; DB 2; Length 542;  
Best Local Similarity 26.9%; Pred. No. 3.4e-08;  
Matches 94; Conservative 68; Mismatches 124; Indels 63; Gaps 16;  
QY 113 KNTTEP---TKKSE-KOLECLLT-HCIKVPMVLDPALPANITLKVPSLYSPHSASDI 167  
DB 14 RRTLEPIVTRQGVARLESSLTPEAQIEDLIF---LRKALNRADIPFLFIRNKNRPV 70  
QY 168 FNVAKPKNPSTNVSVVFDST-----KQVEDAHGLKGNRSQRTWGYLTDTKEVPLV 222  
DB 71 LAINKIKRPAVERALVTACSEPMYAKTIDB-----RGJSPVLVAKGQSQSID-PRIV 123  
QY 223 LMQ-----DLAFLGPPPTFKETNOLKLPENLSKVKLLQLYSEAS 265  
DB 124 RLYRRRIAPGGFRGSRFGVELQWS-----FEET-LIRCPVENSRLRKVLPRK---EVT 174  
QY 266 VALIKLNPK--DQELNKKQTKMNT-----IDGKELTISPALLNDLSAISOSKQ 314  
DB 175 PATIKLYGKWHHTIEGFTPHASVTFDIDLVFVSWVDGSD-----PEFRARRAEMSHVV 230  
QY 315 DE-DISASRPEDNELARYSIRERHAPVWRNIEFVINGQIPSWNLNDNPRVITVTHQDV 373  
DB 231 GEGDADARIQIDELXYARSVMFAPWTRIRFIATDSIPPSWL-ADHPMIVIVPAEDH 289  
QY 374 FRNLHLPTFSPAPRESHHRIEGLSQKFIYLNDDVMFGKDVWPDFFYS 422  
DB 290 FSDRSALPTYNHAVESQLHRIPDLSEHFLYSNDMDFGRPLKASMPFS 338

RESULT 7  
S42434  
Hypothetical protein - Neisseria meningitidis  
C:Species: Neisseria meningitidis  
C>Date: 07-May-1998 #sequence\_revision 15-May-1998 #text\_change 09-Jul-2004  
C:Accession: S42434  
R:Hammerschmidt, S.; Birkholz, C.; Zaehring, U.; Robertson, B.D.; van Putten, J.; Ebel  
Mol. Microbiol. 11, 885-896, 1994  
A:Title: Contribution of genes from the capsule gene complex (cps) to lipooligosaccharid  
A:Reference number: S42430; MUID:94293762; PMID:8022265  
A:Accession: S42434  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-373 <HAM>  
A:Cross-references: UNIPROT:Q51151; EMBL:L09189; NID:g388931; PIDN:AAA63160.1; PID:g3889  
A:Experimental source: strain B1940  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1993

Query Match 4.7%; Score 228.5; DB 2; Length 373;  
Best Local Similarity 31.3%; Pred. No. 9.4e-07;  
Matches 52; Conservative 36; Mismatches 51; Indels 17; Gaps 5;

QY 271 LNNPKDFQBLNKQTKQMTIDGKELTISPAYLLWD-----LSAISQSKQDEDI-----S 319  
DB 43 INADANLASLTENVLPNPFIDVVFTVWNTDKAWEQYRTLPQIDQ-----EDIGUYATD 98  
QY 320 ASRFEDNEELRYSLRSIERHAPWRNIFIVTNGQIPSWNLND-NPRVITVTHQDVFRNLS 378  
DB 99 PARFSNHNELFYVQAVQXFMFWRNIFIVTADQPKWLDENIHSKILINHSQIT-DAK 157  
QY 379 HUPTFSSPAIESHHRIEGLSQKFIYLNDDVMFGKDVWPDFFYS 424  
DB 158 YLPTFNHSHVIEANLYKIPDLSEHFIYFNDVDFVARPLMPNHFENN 203

RESULT 8  
T30201  
Notch homolog protein - sea squirt (Halocynthia roretzi)  
C:Species: Halocynthia roretzi  
C>Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 08-Sep-2002  
C:Accession: T30201  
R:Hori, S.; Saitoh, T.; Matsumoto, M.; Makabe, K.W.; Nishida, H.  
Dev. Genes Evol. 207, 371-380, 1997  
A:Title: Notch homologue from Halocynthia roretzi is preferentially expressed in the c  
A:Reference number: Z20775  
A:Accession: T30201  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-2352 <HOR>  
A:Cross-references: EMBL:AB001327; NID:d1204472; PID:d1026501; PIDN:BAA25571.1  
C:Genetics:  
A:Gene: Notch  
C:Superfamily: notch protein; ankyrin repeat homology; EGF homology

Query Match 3.6%; Score 175.5; DB 2; Length 2352;  
Best Local Similarity 20.3%; Pred. No. 0.034;  
Matches 112; Conservative 70; Mismatches 190; Indels 179; Gaps 26;  
QY 438 CAEGCPGSWIK-DGYCDK--ACNNSACDWDGDCS-GNSGSRVYAGGGTSGVGHWP 493  
DB 1308 CSNTC--ELVKGDI-CDEIGCNVHECFDGDGSLGIS-----ILGCKDKELRCWQY 1360  
QY 494 QFGGGINSVSYNQGG-----ANSLADKFCDCQACNVLS 527  
DB 1361 QDG-----HDEHNNEDCLHDGWDGCVGDCNPKYKCYEYNNYDCQGCNNAN 1413  
QY 528 CGFDAGDCGQDH---FHELYKVILLPNQTHYIIPKGRCLPYFSPAEVAKRGVEGAYSNP 584  
DB 1414 CGWDGADCTIKDYPAQYARGFLVIRMEGITPEEMREQEMIFIEFGKILRSIGIE----- 1466  
QY 585 IIRHASIANKWTIHLIMHSGMNATTIHFNLTFQNT-----NDEEFKQITVEVDTRE-- 637  
DB 1467 IISIEI-----FLESSRRRALLDFDVMESKTRRRATDTSATSEVKVRLNRLCL 1518  
QY 638 -----GPKANSTAQGYENL---VSPITLLPEAILFEDIPKEKRP----- 676  
DB 1519 LNDSCFWEASSAASVVSASASTGLQVDFVSTATSQP-----DETSDPETLVVIL 1571  
QY 677 -----KFKHDVNSTRAQEVKIPL-----VNISLLPK 705  
DB 1572 SSGVGVLLVAVVLRKREKRENTLWLPFGVGVKKRREPIGGDQDLNKKLUSK 1631  
QY 706 DAQLSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQHAKIKKQAIITDETNDLSVAP 765  
DB 1632 SSQDLPNSTPFIHSHWEDGNQKHQSKYLAQDESLSGSLPSYTNQS---KFNLSAL--- 1684  
QY 766 QEKQVHKSI---LPNSL---GVSELRQLRTPAVSVKNGHDQGNQPLDLETTARFRVE 819  
DB 1695 SDREAHDPVRKWTPOHLQAGLSSELL-----PSSPLTENGQDLND---IDI----- 1727  
QY 820 THTQKTIGNVTKEXPPSLVPLESQMTKEK--KITGKEKNSRMEENAHNIG--VTEV 875  
DB 1728 -----RGPDCITPLWASTRGGGIEVVSDEESAEGLSESENIANLLTQ- 1772

QY 876 LLGRKLOHYTD 886  
Db 1773 --GASLSAQTD 1781

RESULT 9  
A49128  
cell-fate determining gene Notch2 protein - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 21-Jan-1994 #sequence\_revision 18-Nov-1994 #text\_change 16-Aug-2004  
C:Accession: A49128  
R:Weinmaster, G.; Roberts, V.J.; Lemke, G.  
Development 116, 931-941, 1992  
A:Title: Notch2: a second mammalian Notch gene.  
A:Reference number: A49128; MUID:93202015; PMID:1295745  
A:Accession: A49128  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-2471 <WEI>  
A:Cross-references: UNIPROT:Q9QW30  
A:Experimental source: Schwann cell  
A:Note: sequence extracted from NCBI backbone (NCBIP:127811)  
C:Superfamily: Notch protein; ankyrin repeat homology; EGF homology  
F:264-295/Domain: EGF homology <EGX1>  
F:799-830/Domain: EGF homology <EGF1>  
F:877-908/Domain: EGF homology <EGX2>  
F:1029-1060/Domain: EGF homology <EGF>  
F:1067-1098/Domain: EGF homology <EGX3>  
F:1153-1184/Domain: EGF homology <EGF3>  
F:1191-1222/Domain: EGF homology <EGX4>  
F:1876-1908/Domain: ankyrin repeat homology <AN1>  
F:1909-1941/Domain: ankyrin repeat homology <AN2>  
F:1943-1975/Domain: ankyrin repeat homology <AN3>  
F:1976-2008/Domain: ankyrin repeat homology <AN4>  
F:2009-2041/Domain: ankyrin repeat homology <AN5>

Query Match 3.4%; Score 168.5; DB 2; Length 2471;  
Best Local Similarity 26.2%; Pred. No. 0.1;  
Matches 44; Conservative 15; Mismatches 58; Indels 51; Gaps 5;

QY 417 PDDFYSHKQKQVLTVPVNCAGCGSIKQGYCDXACNNSACDWDGDCSGSGGR 476  
Db 1404 PPFMGSHCESYAPTSTPTATCLISQYCADKARDGICDEACNSHACQWDGDC-----1455

QY 477 YIAGGGTSGIVGHPW-QFEGGINSVSYCNQGG-----AN 511  
Db 1456 -----LUTMEDPANCSTSLRCWEYINQCDELNCABCLFDFECQNSKTCYD 1506

QY 512 SWLADKF-----CDQACNVLSGDFDAGCGQDHFHELYK-----VILLP 550  
Db 1507 KYCADHFKDNCXGKCNNECGWDGLDCAADQDENLAEGILVIVLLP 1554

RESULT 10  
S06434  
homeotic protein lin-12 precursor - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Jan-1993 #sequence\_revision 29-Jan-1993 #text\_change 16-Aug-2004  
C:Accession: S06434; A24769  
R:Yochem, J.; Weston, K.; Greenwald, I.  
Nature 335, 547-550, 1988  
A:Title: The Caenorhabditis elegans lin-12 gene encodes a transmembrane protein with over 12 extracellular EGF repeats.  
A:Reference number: S06434; MUID:88334747; PMID:3419531  
A:Accession: S06434  
A:Molecule type: DNA  
A:Residues: 1-1429 <YOC>  
A:Cross-references: UNIPROT:P14585; EMBL:M12069; NID:g156357; PIDN:AAA70191.1; PID:g156357  
R:Greenwald, I.  
Cell 43, 583-590, 1985  
A:Reference number: A24769; MUID:86079540; PMID:3000611  
A:Accession: A24769  
A:Molecule type: DNA  
A:Residues: 173-712 <GRE>

C:Genetics:  
A:introns: 50/2; 90/1; 109/1; 172/3; 545/1; 589/2; 632/2; 1273/3; 1389/3  
C:Superfamily: ankyrin repeat homology; EGF homology  
C:Keywords: glycoprotein; transmembrane protein  
F:254-284/Domain: EGF homology <EGF1>  
F:507-540/Domain: EGF homology <EGF>  
F:547-578/Domain: EGF homology <EGF2>  
F:909-931/Domain: transmembrane #status predicted <TM>  
F:1093-1125/Domain: ankyrin repeat homology <AN1>  
F:1206-1238/Domain: ankyrin repeat homology <AN2>  
F:1240-1272/Domain: ankyrin repeat homology <AN3>

Query Match 3.4%; Score 166.5; DB 2; Length 1429;  
Best Local Similarity 20.4%; Pred. No. 0.058;  
Matches 113; Conservative 43; Mismatches 172; Indels 225; Gaps 22;

QY 438 CAEGCPGSIWMDGYCDXACNNSACDWDGDCSGNS---GGSRVYAGGGTSGIVGHPWQ 494  
Db 643 CSERA-----NDGNCADACNVACKFDGDCSGKREPFKCRY-----GNMCADF-- 687

QY 495 FGGINSVSYCNQCCAN-----SWLADKFCDQACNVLS 528  
Db 688 FANGV-----CNQACNNEECYDGMCLPAVVRCPVKIREHCASRFANGICDPECNNGC 742

QY 529 GFDAGDCQDHFHELYKVVLLPNQTHYIIPKGECLPYFSFAEVAKRGVEGAYSDNP1IRH 588  
Db 743 GFDGDC-----DNET-----753

QY 589 ASIANKWKTHLIMHSGNNATTI-HFNLTFTQNTDEEPK-----MOLITVEVD 634  
Db 754 -----NATIITNIRITVQ-MDPKSFQVTGGQSLMEISALRTVTRIQ 794

QY 635 -TREGKLNSTAQGYENLVSPITL---LPEAEILFEDIPKPKRKHHDVNSTERR 689  
Db 795 RDEGLV---FQNGESEDVDRVKNERQULTEHVLSTSR-----KIKSATNIGVVV 846

QY 690 QEEVKIPLVNLISLPKDAQSLNTLDLQLEHGDITLKGYNLS-----KSALLRSFL- 740  
Db 847 YLEVCENDTCKLYKDAQSVVDSISARLAKKIGDSFGIFISEALVAEPKSGNNTGFLS 906

QY 741 -----MNSOHAQKNOAIITDET 758  
Db 907 WNALLIAGCLIVVVLMLGALPCNTRKTRKRMNASVWPPMENEKKNQHCISITSSQ 966

QY 759 NDSLVAPOEKV-----HKSILPNSLVSE-----RLQRLTFPVAIVKNG 799  
Db 967 HSLLEASVDGIKQRNELQHYSLVPNPQGYGNGNDLFGDFNHTNLQIPTPEPEPIKL 1026

QY 800 HDQGNPDLLETARFVE---THQKT---IGGVTKKPPSLIVPLESQMTREKKI 852  
Db 1027 HTEAGSYAITEPITRESVNIIDPRNRTVLHWTASNSAEKSEDLIV-----HEAKECIA 1082

QY 853 TGKEKENSMEEN 865  
Db 1083 AGADVNMDCDEN 1095

RESULT 11  
S18188  
notch protein homolog - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 02-Aug-2002  
C:Accession: S18188  
R:Weinmaster, G.; Roberts, V.J.; Lemke, G.  
Development 113, 199-205, 1991  
A:Title: A homolog of Drosophila Notch expressed during mammalian development.  
A:Reference number: S18188; MUID:92111383; PMID:1764995  
A:Accession: S18188  
A:Molecule type: mRNA  
A:Residues: 1-2531 <WEI>  
A:Cross-references: EMBL:X57405; NID:957634; PID:g57635  
C:Superfamily: notch protein; ankyrin repeat homology; EGF homology  
F:987-1018/Domain: EGF homology <EGF1>

Query Match	3.3%	Score 163;	DB 2;	Length 2531;
Best Local Similarity	20.4%;	Frag. No. 0.23;		
Matches 119; Conservative	68;	Mismatches 204;	Indels 192;	Gaps 29
QY	434	PVPCASGC--PGSWIKDG--CYDKACNNSACDWDGDCSGNSGGG-----	-XII	478
Ddb	1441	PPQIEAECLPFCQVDAGKNKVCNLQCNNHACGWGDGCSLNFNPWKCTOSLCQWYF	1500	
QY	479	AGG-----GGTGSIGVGHPWPFGGINSVSYCNI---QGCAWSWLADKCFDQAQCNVLSC	528	
Ddb	1501	SDGHCDSCQNSAGCLPDGDFCOLTEG-----QCNPPLYDQYCKDHFP-SDGHCDQGCNSAEC	1550	
QY	529	GFDAGDCGDQHPHE-----LYKVILLP-----NCTHYIIPKGELPFYFAEVAERGV	576	

Db 1555 EWDGLDCA-EHVPRLAAGTLVLLPPDOLRNNSPHFL---RELSHVLHTNVVFKRDA 1610  
 QY 577 EGAYSNDPIIRHASTANK-----WKTIIHLI-----MHSGMNTTIHNLTFONT 620  
 Db 1611 QQQQMFPPYHHEELRHPHKRSTVGWATSSLLPGTSGGRRELDPMDIRGSIVY--- 1667  
 QY 621 NDEEFKQITVEVDTRE-----GPKLNTAOGYENLVSPITLLPEAEIL 665  
 Db 1668 -----LEIDNRQCVOSSQCFQSATDVAALFGLASLSLNIPYKIEAV-KSEPV 1716  
 QY 665 FEDIP-----KKEK-----FKP-KRHDVNSTRAQE 691  
 Db 1717 EPLPQLHLMVAAAFVLLFVCGVLLSRKRRRQHGQWPFEGFKVSEASKKRR- 1775  
 QY 692 EVKIFL--VNISLLP-KDAQLSLNTLDLQLEHG--DITLKGYNLSKALLRSLFMSQA 746  
 Db 1776 ---ELGDSVGLKPLKNASDCALMDNDNONGWDEDLTKKFRPEEPVWLPDLSDQTDHR 1831  
 QY 747 KIKNQALITDETNDLSLVAPOEQVHKSLPNSLGLVSELRQLTTPPAVSVKVNGHQQNP 806  
 Db 1832 QWTOQHLDAADLUMGAMAPTPPQ-----GEVDADCMVNVVRGPD--GFT 1873  
 QY 807 PLDLTTARFRVETHTKTIGNTVTKKPPSLI-----VPLESQMTKEKIT----- 853  
 Db 1874 PLMIASCSGGGLT-----GNSBEEDAPAVISDFIYQASLHNQTDRTGETALHLAAR 1927  
 QY 854 -CKEKENSMEENAEHNI-----GVTEVILL 877  
 Db 1928 YSRDRRKRLEASADANIQDNMGRTPLHAASVADAQGVFOILL 1970

## RESULT 13

S42612  
 C:Species: Brachydanio rerio (zebra fish)  
 C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Jul-2004  
 C:Accession: S42612  
 R:Bierkamp, C.; Campos-Ortega, J.A.  
 Mech. Dev. 43, 87-100, 1993  
 A:Title: A zebrafish homologue of the Drosophila neurogenic gene Notch and its pattern c  
 A:Reference number: S42612; MUID:94128602; PMID:8297791  
 A:Accession: S42612  
 A:Status: Preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-2437 <BIE>  
 A:Cross-references: UNIPROT:P46530; EMBL:X69088; NID:q433866; PIDN:CAA48831.1; PID:q4338  
 C:Superfamily: notch protein; ankyrin repeat homology; EGF homology  
 F:755-786/Domain: EGF homology <EGF1>  
 F:1023-1054/Domain: EGF homology <EGF2>  
 F:1185-1216/Domain: EGF homology <EGF3>  
 F:1915-1947/Domain: ankyrin repeat homology <AN1>  
 F:1948-1980/Domain: ankyrin repeat homology <AN2>  
 F:1982-2014/Domain: ankyrin repeat homology <AN3>  
 F:2015-2047/Domain: ankyrin repeat homology <AN4>  
 F:2048-2080/Domain: ankyrin repeat homology <AN5>

Query Match 3.3%; Score 162.5; DB 2; Length 2437;  
 Best Local Similarity 20.7%; Pred. No. 0.24;  
 Matches 119; Conservative 58; Mismatches 201; Indels 197; Gaps 29;  
 QY 421 YSHSKGQ-----KVLVTPVPCAEPCGPGSWIKDGYCDKACNNSACDWGDCGNS 472  
 Db 1427 YSFGSGGQGRDIAPPVEVIRCEIAQC-EGRGNAI-----CDTQCNNHACGWGDCGDC- 1477  
 QY 473 GGSRYIAGGGTGSIGVHPWQ-----FGGINSVSYCNOCGANS----- 512  
 Db 1478 -----SLNFDDPNQNSAALQWRVFDNG-----KDCQCATAGCLYDGFDC 1519  
 QY 513 -----WLADKFCQACNVLSGCFDAGDCGQD-----HFHELYKVLVLP 550  
 Db 1520 QRLGQCNPYDQYCRDHYADGHGDCGNCNAECWGLDCCADDVPOKLAVGSILVWHVP 1579  
 QY 551 -----NQTHYII-----PKGECL--PYFSFAEVAKRGVEGAYSDNPPIR 587

Db 1580 PDELRSSFLRLSSLLHTNVVFRDANGEALIFPYG-----SEHLSK 1626  
 QY 588 HASTANKWKTIIHLMHSGMNTTIHFN-LTFQNTNDEEFKQIT-VEVDTRE----- 637  
 Db 1627 HK--RSDWTDPGQMQBARSLTSFLKPRTRRELDMHEVKGSIYVLEIDNRQCFQSDRC 1684  
 QY 638 -----GPKLNTAOGYENLVSPITLLPEABILPEDIKPKRFPKFRHVDVNSTRAQ 690  
 Db 1685 FQSATDVAALFGLASSG--NLNVF-----YILEVTSSEGGPK-----TG 1723  
 QY 691 BEVKIPLVNLISLLPKDAQLSLNTLDL-----OLEHGDTLKGYNLSKALLRSLFMSQ 744  
 Db 1724 EYIPMFLVLLAL-----AVLAAAGVVVSVSRKRREHGQWFP-----EGFKVNEP 1769  
 QY 745 HAKIKNQALITDETNDLSLVAPOEQVHKSLPNSL--GVSELRQLTTPPAVS-KVNG- 799  
 Db 1770 KKKREP-----GEDSVGLKPLKNSSSLMDQLSEMAEDDTNKFRFEGQSILEMSQ 1824  
 QY 800 --HQGQNPPLDLETTARFRVETHTKTIGNVTK-----KPPSLIVPLESQMTKEKK 851  
 Db 1825 LDHRQWTQOHLD-----AADRLNSMNAPTPQOQIENDCMVNVVRGPDGFTFL---MIASCS 1878  
 QY 852 ITGKEKENSMEENAEHNI--GVTEVILLGRKLQHYTD 886  
 Db 1879 GGGLENGENGEAEDPSADVTITDFIYHGANLHNQTD 1913

## RESULT 14

S78549  
 C:Species: Homo sapiens (man)  
 C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 09-Jul-2004  
 C:Accession: S78549; S71825  
 R:Joutel, A.; Tournier-Lasserre, E.  
 Submitted to the EMBL Data Library, April 1997  
 A:Reference number: S78549  
 A:Accession: S78549  
 A:Molecule type: mRNA  
 A:Residues: 1-2321 <JOU1>  
 A:Cross-references: UNIPROT:Q9UM47; EMBL:U97669; NID:q2668591; PIDN:AAB91371.1; PID:q266  
 X:Joutel, A.; Corpechot, C.; Ducros, A.; Vahedi, K.; Chabrier, H.; Mouton, P.; Alamowit  
 X: M.M.; Weissenbach, J.; Bach, J.F.; Bousser, M.G.; Tournier-Lasserre, E.  
 Nature 383, 707-710, 1996  
 A:Title: Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke  
 A:Reference number: S71825; MUID:97032728; PMID:8878478  
 A:Accession: S71825  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 67-113;138-194;268-333; 'G', 335-346;536-613;716-765;1240-1279;1815-1888 <JOU2  
 A:Cross-references: EMBL:U97669  
 C:Genetics:  
 A:Gene: notch3  
 A:Map position: 19p13.1  
 C:Function:

A:Description: may be involved in pathogenesis of CADASIL, causing a type of stroke and  
 C:Superfamily: notch protein; ankyrin repeat homology; EGF homology  
 C:Keywords: tandem repeat; transmembrane protein  
 F:123-155/Domain: EGF homology <EGX1>  
 F:162-194/Domain: EGF homology <EGF1>  
 F:240-271/Domain: EGF homology <EGX2>  
 F:318-349/Domain: EGF homology <EGF>  
 F:473-504/Domain: EGF homology <EGX3>  
 F:853-884/Domain: EGF homology <EGF3>  
 F:928-959/Domain: EGF homology <EGX4>  
 F:1838-1876/Domain: ankyrin repeat homology <AN1>  
 F:1871-1903/Domain: ankyrin repeat homology <AN2>  
 F:1905-1937/Domain: ankyrin repeat homology <AN3>  
 F:1938-1970/Domain: ankyrin repeat homology <AN4>  
 F:1971-2003/Domain: ankyrin repeat homology <AN5>

## Query Match

Best Local Similarity 3.1%; Score 154.5; DB 2; Length 2321;  
 Pred. No. 0.69;

```
Matches 38; Conservative 14; Mismatches 37; Indels 65; Gaps 6;
QY 433 WVPVNC-----AEGCPGSWK-----DGYCDKACNNACDWDGDCSGNSGGSKY 477
Db 1367 WTGPRCEAPAAPEVSEPRCPRAACQAKRGQRCDRECNPGCGWDGGDC-----1417
QY 478 IAGGGGTSGVGHWPQFGGIN-----SVSYCNQGCAN-----511
Db 1418 -----SLVGDPRQCEALQCWLRFNNRCDPACSPACLYDNFDCHAGGRERTCNP 1469
QY 512 ----SWLADKF-----CDQACNVLSGCFDAGDCQD 538
Db 1470 VZEKYCADHFADGRCDQGCNTEECGWDGLDCASE 1503

RESULT 15
notch4 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C:Accession: T09059
R:Rowen L.; Mahairas, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Lasky, S.; Loretz, C.; So
submitted to the EMBL Data Library, October 1997
A:Description: Sequence of the mouse major histocompatibility locus class III region.
A:Reference number: Z16543
A:Accession: T09059
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1964 <ROW>
A:Cross-references: UNIPROT:P31695; EMBL:AF030001; NID:G2564945; PID:G2564947
C:Genetics:
A:Gene: notch4
A:Map position: 17
A:Introns: 22/1; 49/2; 148/1; 264/1; 305/1; 384/1; 436/1; 501/1; 539/1; 577/1; 618/1; 67
1679/3; 1729/1; 1761/3
C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
C:Keywords: receptor; signal transduction
F:514-545/Domain: EGF homology <EGF>

Query Match 3.1%; Score 152; DB 2; Length 1964;
Best Local Similarity 28.1%; Pred. No. 0.76;
Matches 41; Conservative 11; Mismatches 32; Indels 62; Gaps 8;
QY 434 WVPVNC-----AEGCPGSWKDGYCDKACNNACDWDGDCSGNSGGSKYIAGGGGTSGV 489
Db 1162 PGPRCQRPAGSGCEGRG-GDGTCDAGCGSGPGGDWDGGDC-----SLGV 1203
QY 490 GHPWQFGGGINSVYC-----NQCA-----NSWLADKF- 518
Db 1204 PDFWK---GCPFHQCWLLFRDGRCHPCQDSECLFDGYDCEIPLTCIPAYDQYCRDHFH 1260
QY 519 ---CDQACNVLSGCFDAGDC---QQD 538
Db 1261 NGHCEKGCNNAECGWDGDCRPEGED 1286

Search completed: November 21, 2004, 12:56:00
Job time : 30.3503 secs
```

***This Page Blank (uspto)***



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2004, 12:46:58 ; Search time 133.057 Seconds  
(without alignments)  
4012.911 Million cell updates/sec

Title: US-10-023-888-4

Perfect score: 4907

Sequence: 1 MLFKLQRQTYTCLSHRYGL.....SLKTQLAYFTDSKNTGRQLK 928

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 02.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4473	91.2	847	Q96N13	Q96N13 homo sapien
2	3314	67.5	950	Q9UL12	Q9UL12 homo sapien
3	2565.5	52.3	921	Q6P9F5	Q6P9F5 mus musculus
4	2565.5	52.3	921	AAH60638	AAH60638 mus musculus
5	2495	50.8	490	Q6IPW5	Q6IPW5 homo sapien
6	2495	50.8	490	AAH71687	AAH71687 homo sapien
7	837	17.1	490	Q86TQ2	Q86TQ2 homo sapien
8	734.5	15.0	384	Q61340	Q61340 mus musculus
9	488	9.9	666	Q9V553	Q9V553 drosophila
10	488	9.9	666	Q8SX14	Q8SX14 drosophila
11	486.5	9.9	717	Q7Q098	Q7Q098 anopheles g
12	360	7.3	71	Q9BUA5	Q9BUA5 homo sapien
13	329	6.7	602	Q9L112	Q9L112 streptomyc
14	324.5	6.6	942	Q9L114	Q9L114 streptomyc
15	322.5	6.5	586	Q69853	Q69853 streptomyc
16	320.5	6.5	486	Q7X4S1	Q7X4S1 neisseria m
17	314	6.4	541	Q69851	Q69851 streptomyc
18	287	5.8	238	Q8GPD3	Q8GPD3 streptococc
19	287	5.8	238	Q9EVX1	Q9EVX1 streptococc
20	266	5.4	545	Q83U59	Q83U59 neisseria m
21	266	5.4	545	Q84D00	Q84D00 neisseria m
22	266	5.4	545	Q9JW88	Q9JW88 neisseria m
23	265	5.4	545	Q68215	Q68215 neisseria m
24	260.5	5.3	442	Q86TW6	Q86TW6 dictyostell
25	259	5.3	545	Q84CZ9	Q84CZ9 neisseria m
26	258	5.3	333	Q8GP72	Q8GP72 streptococc
27	258	5.3	483	Q86X00	Q86X00 dictyostell
28	256.5	5.2	364	Q714U9	Q714U9 haemophilus
29	256.5	5.2	364	AAQ12659	AAQ12659 haemophil
30	256	5.2	532	Q06628	Q06628 mycobacteri
31	255	5.2	532	Q7U184	Q7U184 mycobacteri

## ALIGNMENTS

RESULT 1

Q96N13

ID Q96N13

AC Q96N13 PRELIMINARY; PRT; 847 AA.

DT 01-DEC-2001 (Tremblrel. 19, Created)

DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)

DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)

DE Hypothetical protein FLJ31575.

OS Homo sapiens (human)

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

CX NCBI\_TaxID=9606;

[1]

RP SEQUENCE FROM N.A.

RX PubMed=14702039;

RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,

Wakanatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,

Sekine M., Ohtsubashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,

Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,

Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,

Sudo H., Hosioki T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,

Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,

Abe K., Kamiya K., Katsura N., Sato K., Tanikawa M., Yamazaki M.,

Tanai H., Kimata M., Watanabe M., Hiraoa S., Chiba Y., Ishida S.,

Ono Y., Takiguchi S., Watanabe S., Yosida M., Hctuta T., Kusano J.,

Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,

Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,

Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,

Yoshikawa Y., Matsunawa H., Ichihara T., Shionata N., Sano S.,

Moriya S., Momiya H., Satoh N., Takami S., Terashima Y., Suzuki O.,

Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,

Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,

Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,

Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara Y.,

Ono T., Yamada K., Fujii Y., Ozaki K., Hiroo M., Ohmori Y.,

Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,

Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,

Togashi K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,

Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,

Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,

Okumura K., Nagase T., Nomura N., Kikuchi Y., Masuho Y., Yanashita R.,

"Complete sequencing and characterization of 21,243 full-length human

CDNAs";

Nat. Genet. 36:40-45 (2004).

EMBL; AK056137; BAB71102.1;

GO; GO:0016020; C:membrane; IEA.

GO; GO:0030154; P:cell differentiation; IEA.

InterPro; IPR010506; P:cell binding.

InterPro; IPR000800; DMAP-binding.

Pfam; PF08464; DMAP-binding; 1.

Pfam; PF00066; Notch; 1.

SMART; SM00004; NL; 2.

SMART; SM00004; NL; 2.

FT NON TER 847 847  
SQ SEQUENCE 847 AA; 95165 MW; 0A115015824733CS CRC64;

Query Match 91.2%; Score 4473; DB 2; Length 847;  
Best Local Similarity 99.8%; Pred. No. 1.1e-269;  
Matches 845; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLFKLQRTYTCLSHRGLYVCFGLGVVTVISAFQGEVVLWSDQYHVLFDSDYDNI 60  
DB 1 MLFKLQRTYTCLSHRGLYVCFGLGVVTVISAFQGEVVLWSDQYHVLFDSDYDNI 60

QY 61 AGSKFQNRCLCLPMPIDVVTYVWNGTDLLELLEKQVREOMBEQKAMREILGNKNTTEPTK 120  
DB 61 AGSKFQNRCLCLPMPIDVVTYVWNGTDLLELLEKQVREOMBEQKAMREILGNKNTTEPTK 120

QY 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKVPSLPSPFHSASDIFNVAKPNPSTNV 180  
DB 121 KSEKQLECLLTHCIKVPMLVLDPALPANITLKVPSLPSPFHSASDIFNVAKPNPSTNV 180

QY 181 SVVFDSTKVEDAHSGLLKNSQRTVWRGYLTDDKEVPGVLVLMODLAFLSGPPPTFKET 240  
DB 181 SVVFDSTKVEDAHSGLLKNSQRTVWRGYLTDDKEVPGVLVLMODLAFLSGPPPTFKET 240

QY 241 NQKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNTIDGKELTISPA 300  
DB 241 NQKTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKQTKKNTIDGKELTISPA 300

QY 301 YLLNLSAISQSKQEDISASFPEDNEELRYSLRSHAPVWRNPIFVINGQIPSWLNL 360  
DB 301 YLLNLSAISQSKQEDISASFPEDNEELRYSLRSHAPVWRNPIFVINGQIPSWLNL 360

QY 361 DNPRVTIVTHQDVFNLSHLPTFSSPAIESHRIEGLSQFIYLNDDVNFQKVPDDF 420  
DB 361 DNPRVTIVTHQDVFNLSHLPTFSSPAIESHRIEGLSQFIYLNDDVNFQKVPDDF 420

QY 421 YSHSGQKQYLTWPVPCNAECGPGSWIKDGYCDKACNNSACDWDGDCSGNSGGSRYIAG 480  
DB 421 YSHSGQKQYLTWPVPCNAECGPGSWIKDGYCDKACNNSACDWDGDCSGNSGGSRYIAG 480

QY 481 GGGTSGIYGHPWPGFGGGINSVYCNQGCANSWLADKFCQACNVLSGCGFDAGDCQDHF 540  
DB 481 GGGTSGIYGHPWPGFGGGINSVYCNQGCANSWLADKFCQACNVLSGCGFDAGDCQDHF 540

QY 541 HELYKVIILLPQTHYIIPKGECLPYFSAEVAKRGVEGAYSDNPIIRHASIANKWTIHL 600  
DB 541 HELYKVIILLPQTHYIIPKGECLPYFSAEVAKRGVEGAYSDNPIIRHASIANKWTIHL 600

QY 601 IWHSGMNTTIFNLTFONTNDEEFKQITVEVDTRGPKLNTAQGYENLVSPITLLP 660  
DB 601 IWHSGMNTTIFNLTFONTNDEEFKQITVEVDTRGPKLNTAQGYENLVSPITLLP 660

QY 661 EAEILFEDIPKPKRKHVDVNSTRAAEVKIPLVNISSLPKQAQLSLNTLDLQLEH 720  
DB 661 EAEILFEDIPKPKRKHVDVNSTRAAEVKIPLVNISSLPKQAQLSLNTLDLQLEH 720

QY 721 GDITLKGYNLSKALLRSLMNSOHAQKNOAIITDETNDLSVAPQKOVHKSILNSLG 780  
DB 721 GDITLKGYNLSKALLRSLMNSOHAQKNOAIITDETNDLSVAPQKOVHKSILNSLG 780

QY 781 VSELRQRTFFPAVSKVNGHDQGNPDLLETTARFRVETHQTKTIGGNTVTKPKPSLIV 840  
DB 781 VSELRQRTFFPAVSKVNGHDQGNPDLLETTARFRVETHQTKTIGGNTVTKPKPSLIV 840

QY 841 PLESQMT 847  
DB 841 PLESQMT 847

RESULT 2  
Q9ULL2  
ID Q9ULL2  
AC Q9ULL2;  
DT 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE KIAA1208 protein (Fragment).  
GN Name=KIAA1208;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=brain;  
RX MEDLINE=20039619; Pubmed=10574462;  
RA Nagase T., Ishikawa K., Kikuno R., Hirose M., Nomura N., Ohara O.,  
RT "Prediction of the coding sequences of unidentified human genes. XV.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 6:337-345 (1999).  
DR EMBL; AB033034; BA866522.2; -;  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005509; F:calcium ion binding; IEA.  
DR GO; GO:0030154; P:cell differentiation; IEA.  
DR InterPro; IPR010506; DMAP binding.  
DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR008000; Notch region.  
DR Pfam; PF06464; DMAP binding; 1.  
DR Pfam; PF00036; ehand; 1.  
DR Pfam; PF00066; Notch; 1.  
DR SMART; SM00004; NL; 2.  
DR PROSITE; PS00018; EF\_HAND; UNKNOWN\_1.  
FT NON TER 1 1  
SQ SEQUENCE 950 AA; 108785 MW; F653BDPA6EAC7503 CRC64;

Query Match 67.5%; Score 3314; DB 2; Length 950;  
Best Local Similarity 99.8%; Pred. No. 1.9e-197;  
Matches 621; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 307 SAISQSKQDDEISASFPEDNEELRYSLRSHAPVWRNPIFVINGQIPSWLNLNPRVT 366  
DB 1 SAISQSKQDDEISASFPEDNEELRYSLRSHAPVWRNPIFVINGQIPSWLNLNPRVT 60

QY 367 IVTHQDVFNLSHLPTFSSPAIESHRIEGLSQFIYLNDDVNFQKVPDDFVSHSGK 426  
DB 61 IVTHQDVFNLSHLPTFSSPAIESHRIEGLSQFIYLNDDVNFQKVPDDFVSHSGK 120

QY 427 QKYLTPVPCNAECGPGSWIKDGYCDKACNNSACDWDGDCSGNSGGSRYIAGGGTGS 486  
DB 121 QKYLTPVPCNAECGPGSWIKDGYCDKACNNSACDWDGDCSGNSGGSRYIAGGGTGS 180

QY 487 IGVGHPWPGFGGGINSVYCNQGCANSWLADKFCQACNVLSGCGFDAGDCQDHFELKY 546  
DB 181 IGVGHPWPGFGGGINSVYCNQGCANSWLADKFCQACNVLSGCGFDAGDCQDHFELKY 240

QY 547 ILLPNQTHYIIPKGECLPYFSAEVAKRGVEGAYSDNPIIRHASIANKWTIHLHSGM 606  
DB 241 ILLPNQTHYIIPKGECLPYFSAEVAKRGVEGAYSDNPIIRHASIANKWTIHLHSGM 300

QY 607 NATTIHNLTFONTNDEEFKQITVEVDTRGPKLNTAQGYENLVSPITLLPBAEILF 666  
DB 301 NATTIHNLTFONTNDEEFKQITVEVDTRGPKLNTAQGYENLVSPITLLPBAEILF 360

QY 667 EDIPKPKRKHVDVNSTRAAEVKIPLVNISSLPKQAQLSLNTLDLQLEHGDITLK 726  
DB 361 EDIPKPKRKHVDVNSTRAAEVKIPLVNISSLPKQAQLSLNTLDLQLEHGDITLK 420

QY 727 GYNLSKALLRSLMNSOHAQKNOAIITDETNDLSVAPQKOVHKSILNSLGVSERLQ 786  
DB 421 GYNLSKALLRSLMNSOHAQKNOAIITDETNDLSVAPQKOVHKSILNSLGVSERLQ 480

QY 787 RLTFPAVSKVNGHDQGNPDLLETTARFRVETHQTKTIGGNTVTKPKPSLIVPLESQM 846  
DB 481 RLTFPAVSKVNGHDQGNPDLLETTARFRVETHQTKTIGGNTVTKPKPSLIVPLESQM 540

QY 847 TREKKTIGKENSRMEENAHNIGVTEVLLGRKLQHYTDSYLGFLPWEKKKYFQDLDE 906

```

Db 541 TKEKTKCKENSMENAEHNGVTEVLLGRKLQHYTSYLGFLPWKKYFQDLDLE 600
QY 907 BESLKTQLAYTDSKNTGRQK 928
Db 601 BESLKTQLAYTDSKNTGRQK 622

RESULT 3
Q6P9R5 PRELIMINARY; PRT; 921 AA.
AC Q6P9R5;
DT 05-JUL-2004 (TREMELrel. 27, Created)
DT 05-JUL-2004 (TREMELrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMELrel. 27, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
FN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Frange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Souffard G.G.,
RA Blakesley R.M., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalusz D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences".
RL Proc. Natl. Acad. Sci. U.S.A. 99:15899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC060638; AAH60638.1; -.
DR InterPro; IPR002048; EF-Hand.
DR InterPro; IPR010983; EF-Hand.
DR Pfam; PF00036; eFhand; 1.
DR Pfam; PF00066; Notch; 1.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 921 AA; 104970 MW; 12879CDE978D87BB CRC64;

Query Match 52.3%; Score 2565.5; DB 2; Length 921;
Best Local Similarity 61.4%; Pred. No. 7.6e-151;
Matches 505; Conservative 35; Mismatches 53; Indels 229; Gaps 6;

QY 107 MREILGNNTPTKSKSQECLLTHCIVKPMVLDPALPANTLKDVPSPSFSASD 166
Db 1 MREILGNNTPTKSKSQECLLTHCIVKPMVLDPALPANTLKDVPSPSFSASD 60
QY 167 IFNVAKPKNSTNVVVVFDSTKDVDAHSGLLKGNRSQTVWVGYLTTDXEVPGLVLMQD 226
Db 61 MFNVAKPKNSTNVVVVFDSTKDVDAHAGPFGKSGQVWVAYLTTDXEAGLVLMQD 120
QY 227 LAFLSGFPPTFKETNQLKTLPENLSSKVKLLQLYSEASVALLKLNPNKDFBELNKQTK 286

```

```

Db 121 LAFLSGFPPTFKETNQLKTLPEKSSKILRLISEASVALLKLNPNKDFBELNKQTK 180
QY 287 NMTIDGKELTISPAYLLWDLSSAISQSKQDEDISAGRFEDNEELRYSLRSIERHAPVVRNI 346
Db 181 NMTIDGKELTISPAYLLWDLSSAISQSKQDEDISAGRFEDNEELRYSLRSIERHAPVVRNI 240
QY 347 FIVTNGOIPSWNLNDNPRVTIIVTHODVFENLHSLTFSSPAIESHRIEGLSOKFIYLN 406
Db 241 FIVTNGOIPSWNLNDNPRVTIIVTHODIFONLSHLTFSSPAIESHRIEGLSOKFIYLN 300
QY 407 DDMFGKDWPDFFYSHSGKQVYLTPVPNCAEGCPGKSWIKDGYCDKACNNSACDWDGG 466
Db 301 DDMFGKDWPDFFYSHSGKQVYLTPVPNCAEGCPGKSWIKDGYCDKACNNSACDWDGG 360
QY 467 DCSNGSGSRYTAGGGTGSIGVHPWQFGGINSVYCNQCSANSWLADKFCDDACNVL 526
Db 361 DCSNGTAGNRFVAGGGGTGIGAGQWQFGGINTISYCNQCSANSWLADKFCDDACNVL 420
QY 527 SCGFADAGCGQDHFHLYKVILLPNQTHVIIIPKGCPLPYFSPAEVAKRGVEGAYSNDPII 586
Db 421 SCGFADAGCGQDHFHLYKVILLPNQTHVVPKGYLSYFSFAG----- 464
QY 587 RHASIANKWKTHILMHSGNNATTIHPNLTFTONTDEEFKMOITVEVDTREGPKLNSTAQ 646
Db 465 ----- 464
QY 647 KGVNLVSPITLLPEAEILFEDIPEKRPFKRHDVNSTREAOBEVKIPLVNISLLPKD 706
Db 465 ----- 464
QY 707 AQLSLNTLDLQLEHGDIITLKGYNSKALLRSLFNLNSHAKIKNOAIITDETNDLSVAPQ 766
Db 465 -----EH----- 466
QY 767 EKQVHKSILPNSLGVSELRQRTFPAVSVMKNGHCGQGNPPDLDTTARFVETHTQKTI 826
Db 467 -----RSEWTAPAEVTYVKGROHNLNPPPVLETNARL-----AQPTL 504
QY 827 GGNVTKEKPPSLIVPLESOMTEKKITGKEKNSMEENAEHNGVTEVLLGRKLQHYTD 886
Db 505 GVTVSKENLSPILVPPESHLP-----KEESDRAEGNA---VPVKELVPERRIQ---Q 551
QY 887 SYLGFLPWKKYFQDLDDEESLKTQLAYTDSKNTGRQK 928
Db 552 NYPGFLPWKKYFQDLDDEESLKTQLAYTDSKNTGRQK 593

RESULT 4
AAH60638 PRELIMINARY; PRT; 921 AA.
AC AAH60638;
DT 02-VAR-2004 (TREMELrel. 27, Created)
DT 02-VAR-2004 (TREMELrel. 27, Last sequence update)
DT 02-VAR-2004 (TREMELrel. 27, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
FN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Frange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

```

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CS7BL/6; TISSUE=Brain;  
RA Strausberg R.;  
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.  
KW EMBL; BC060638; AAH60638.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 921 AA; 104970 MW; 12879CDE978D87BB CRC64;  
  
Query Match 52.3%; Score 2565.5; DB 2; Length 921;  
Best Local Similarity 61.4%; Pred. No. 7.6e-151;  
Matches 505; Conservative 35; Mismatches 53; Indels 229; Gaps 6;  
  
QY 107 MREILGKNTTPEPKKSEKQLECLLTHCIKVPMLVLDLPALPANITLKDVPSPFHSASD 166  
DB 1 MREILGKNTTPEPKKSEKQLECLLTHCIKVPMLVLDLPALPANITLKDVPSPFHSASD 60  
  
QY 167 IFNVAKPNSSTNVSVVVFSTKDVDEAHSGLLKGNRSQRTVRYGILTTKEVPGVLVMD 226  
DB 61 MFNVAKPNSSTNVSVVVFSTKDVDEAHSGLLKGNRSQRTVRYGILTTKEVPGVLVMD 120  
  
QY 227 LAFISGPPPTFKETNQLTKLLENLSKVKLLQYSEASVALLKLNPKDFQBLNKTQKK 286  
DB 121 LAFISGPPPTFKETNQLTKLLENLSKVKLLQYSEASVALLKLNPKDFQBLNKTQKK 180  
  
QY 287 NMTIDGKELTSPAYLWDLISAIQSQKQDEISASRFENEELRSLRISIERHAPVRNI 346  
DB 181 NMTIDGKELTSPAYLWDLISAIQSQKQDEISASRFENEELRSLRISIERHAPVRNI 240  
  
QY 347 FIVTNGQIPSWLMDNPRVTIVTHQVFNRLSHLPTFSSPAISHTRIEGSLQKFIYN 406  
DB 241 FIVTNGQIPSWLMDNPRVTIVTHQVFNRLSHLPTFSSPAISHTRIEGSLQKFIYN 300  
  
QY 407 DDVNFGRDVPDDFYSHSGKQVYLTWVPVNCAGCGPGSWIKDGYCDKACNNSACDWDG 466  
DB 301 DDVNFGRDVPDDFYSHSGKQVYLTWVPVNCAGCGPGSWIKDGYCDKACNNSACDWDG 360  
  
QY 467 DCSNSGSGRYIAGGGTSGVGHFPHQFGGINSYVNCAGNSWLADKCDQACNYL 526  
DB 361 DCSNTAGNFVAGGGTGNIGAGHQHGFQGGINTISYVNCAGNSWLADKCDQACNYL 420  
  
QY 527 SCGFDAGCGQDHFHELYKVLIPNTHYIIPKGECLIPFSPFAEAKRGVEGAYSNDPII 586  
DB 421 SCGFDAGCGQDHFHELYKVLIPNTHYIIPKGECLIPFSPFAEAKRGVEGAYSNDPII 464  
  
QY 587 RHASIANKWKTIHLIMSGNNAATTIHNLTFTQNTNDEEPKQQTVEVDREGKLNSTAQ 646  
DB 465 -----EH----- 464  
  
QY 647 KGVENLYSPITLLPEABILFEDIPKRPFKRHDVNSTRAQEEVKIPLVNISSLPKD 706  
DB 465 ----- 464  
  
QY 707 AQLSANTLDLQLEHGDITLKGYNLSKALLRSFLMNSQHAKIKNQAIITDETNDLSLVAQ 766  
DB 465 ----- 466  
  
QY 767 EKQVHKSILPNSLGVSERLQRTFPVSVKNGHDQCNPPDLTTTARFRVETHQKTI 826  
DB 467 -----RSEWTAPETVTVKGRDHALNPPVLETNARL-----AQPTL 504  
  
QY 827 GGNVTKSKPSPVLPVLESQNTKEKKTITGKSKENSRMEENAEHIGVTEVLLGRKLQHYTD 886

DB 505 GVTYSKENLSPLIVPSSHLP-----KEESDRAGNA---VPVKELVPGRELQ---Q 551  
QY 887 SYLGFLPWEKKKYPQDILLDEESLKTQLAYFTDSKNTGROLK 928  
DB 552 NYPGFLPWEKKKYPQDILLDEESLKTQLAYFTDSKNTGROLK 593  
  
RESULT 5  
QIPWS PRELIMINARY; PRT; 490 AA.  
AC Q61PW5;  
DT 05-JUL-2004 (TREMBlrel. 27, Created)  
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
DE MG4170 protein.  
GN Name=MG4170;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_taxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stopleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Strausberg R.;  
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC071687; AAH71687.1; -  
DR InterPro; IPR000800; Notch\_region.  
DR Pfam; PF00066; Notch; 1.  
DR SMART; SM00004; NL; 1.  
SQ SEQUENCE 490 AA; 55982 MW; 9223139711337ED0 CRC64;  
  
Query Match 50.8%; Score 2495; DB 2; Length 490;  
Best Local Similarity 59.8%; Pred. No. 7.2e-147;  
Matches 469; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MLFLLQRTQYTCLSHRYGLVCFVGVVTVISAFQFGEVLEWSDQYHVFDSYRNI 60  
DB 1 MLFLLQRTQYTCLSHRYGLVCFVGVVTVISAFQFGEVLEWSDQYHVFDSYRNI 60  
  
QY 61 AKGSPQNRCLPMPIDVYVTVWNGTDLLELKELOVREOMEKAREILKNTTPTK 120  
DB 61 AKGSPQNRCLPMPIDVYVTVWNGTDLLELKELOVREOMEKAREILKNTTPTK 120  
  
QY 121 KSEKQLECLLTHCIKVPMLVLDLPALPANITLKVPSLPSPFHSASDIFNVAKPNSSTNV 180  
DB 121 KSEKQLECLLTHCIKVPMLVLDLPALPANITLKVPSLPSPFHSASDIFNVAKPNSSTNV 180  
  
QY 181 SVVFDSTKDVDEAHSGLLKGNRSQRTVRYGILTTKEVPGVLVMDLAFSLGPPPTKET 240

Db	181	SVVVFSTKDVEDAHSGLLKGNRSOTVWRGYLTDTDEKEVPGGLVLMQDLAFUSGPPPTFKET	240
Qy	241	NQLKTKLPENLSKVKLLQLYSEASVALLKLNPKDFQELNKQTKQNTIDGKELTISPA	300
Db	241	NQLKTKLPENLSKVKLLQLYSEASVALLKLNPKDFQELNKQTKQNTIDGKELTISPA	300
Qy	301	YLLWDLISAIQSQKQEDISASPFDENEELRYSLSIERHAPVWNTPIVTVNGOIPSWML	360
Db	301	YLLWDLISAIQSQKQEDISASPFDENEELRYSLSIERHAPVWNTPIVTVNGOIPSWML	360
Qy	361	DNPRVTIVTHQDVFNRLSHLPTFSSPAIESHIHRIEGLSKFYFLNDVVMFGKQVWPDFD	420
Db	361	DNPRVTIVTHQDVFNRLSHLPTFSSPAIESHIHRIEGLSKFYFLNDVVMFGKQVWPDFD	420
Qy	421	YSHSGOKVYLTWPVPCNAEGCPGSWIKDGYCDKACNNSACDWDGDCSG	470
Db	421	YSHSGOKVYLTWPVPCNAEGCPGSWIKDGYCDKACNNSACDWDGDCSG	470
RESULT 6			
AAH71687			
ID	AAH71687	PRELIMINARY;	PRT; 490 AA.
AC	AAH71687;		
DT	01-JUN-2004	(TREMBLrel. 27, Created)	
DT	01-JUN-2004	(TREMBLrel. 27, Last sequence update)	
DT	01-JUN-2004	(TREMBLrel. 27, Last annotation update)	
DE	Hypothetical protein.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
QC	NCBI_TaxID:9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Brain;		
RX	MEDLINE=223388257; PubMed=12477932;		
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,		
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,		
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,		
RA	Hopkins R.F., Jordan H., Moore T.I., Max S.I., Wang J., Hsieh F.,		
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,		
RA	Diatchenko L., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,		
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Pange C.,		
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,		
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,		
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.D., Hulyk S.W.,		
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,		
RA	Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,		
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,		
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,		
RA	Rodriguez A.C., Grinchwood J., Schmutz J., Myers R.M., Butterfield Y.S.,		
RA	Krzyszynski M.I., Skalska U., Smalusz D.E., Schnerch A., Schein J.E.,		
RA	Jones S.J., Marra M.A.;		
RT	"Generation and initial analysis of more than 15,000 full-length human		
RT	and mouse cDNA sequences."		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).		

```

RESULT 8
Q61340 PRELIMINARY; PRT; 384 AA.
AC Q61340;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Basic domain/leucine zipper transcription factor (fragment).
GN Name:Maif;
CS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95094266; PubMed=8001130;
RA Cordes S.P., Barsh G.S.;
RT "The mouse segmentation gene kr encodes a novel basic domain-leucine
RT zipper transcription factor.";
RL Cell 79:1023-1034(1994).
DR EMBL; L36434; AAA65688.1; -.
DR PIR; I49528; I49528.
DR MGD; MGI:104555; Maif.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0005667; C:transcription factor complex; IDA.
DR GO; GO:0003677; F:DNA binding; IDA.
DR GO; GO:0005515; F:protein binding; IDA.
DR GO; GO:0045344; P:positive regulation of transcription from P. . .; IDA.
DR GO; GO:0007585; P:respiratory gaseous exchange; IMP.
DR GO; GO:0007379; P:segment specification; IMP.
DR InterPro; IPR010506; DMAP binding.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF06464; DMAP binding; 1.
DR Pfam; PF00036; efhnd; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
FT NON TER
SQ SEQUENCE 384 AA; 4375 MW; 51F473C6807A7B55 CRC64;

Query Match 15.0%; Score 734.5; DB 2; Length 384;
Best Local Similarity 57.9%; Pred. No. 2.2e-37;
Matches 163; Conservative 28; Mismatches 71; Indels 21; Gaps 6;

Qy 644 TAQGYENLVSPITLLPEABILEDPKFKRPFKFKHVNSTRPAQEEVKIPLVNISLL 703
Db 3 TTQKAYESLVSPVTPLPQADVPPEDVPKRRFPKRRHVDVATGRFPQEEVKIPLVNISLL 62

Qy 704 PKDAOLSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSOHAKIKNOAILTDETNDSLV 763
Db 63 PKAQVRLSLNDLQLEHGDITLKGYNLSKALLRSFLMNSLDTKIIPQA-RIDETKGNLE 121

Qy 764 APQEKQVHKSLTLPNSLGVSRQLRTPPAVSVKVNCHQOQNPPDLDTTARFRVETHQ 823
Db 122 VPQENPSHRR--PHGFAGEHRSERWTAPASTVTVKGRDHALNPPVLETNARL-----AQ 174

Qy 824 KTIQNTVKKPSLIIVPLESQMTKEKIKTKGKENSMEENAHIGVTVELLGRKLQH 883
Db 175 PTLGVTVSKENLSPLIIVPPESHLP-----KSEEDRAEGNA---VPVKELVPGRCRSR 224

Qy 884 YTDSYLGFLPEWKKKYFQDLIDDEESLKTOLAYFTDSKNTGRQLK 928
Db 225 IIQA---FCGKKKKYQDLIDDAEESLKTOLAYFTDSKNTGRQLK 266

RESULT 9
Q9V553 PRELIMINARY; PRT; 666 AA.
AC Q9V553;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

```

```

DE GN ORFNames=CG8027;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.B.D.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abail J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foaier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacieb J.M., M.G.,
RA Palazzolo M., Pittman G.S., Pan S., Poillard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen T.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstock G.M., Weissbach J., Ye J.,
RA Yeh R.P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2183-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celnikier S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacieb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
RT melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celnikier S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,

```



RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Bettencourt B.R., Celiniker S.E., de Grey A.D., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
RT systematic review";  
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083 (2002).  
RN [5]  
RP SEQUENCE FROM N.A.  
RG FLYBASE;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A.  
RG FLYBASE;  
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE003834; AAF58967.2; -;  
DR DR IntAct; Q9V533; -;  
DR FLYBASE; FBgn0033392; CG8027.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0030154; P:cell differentiation; IEA.  
DR InterPro; IPR000800; Notch\_region.  
DR Pfam; PF00066; Notch; 1.  
SQ SEQUENCE 666 AA; 77731 MW; EA232EC57C754FF6 CRC64;  
  
Query Match 9.9%; Score 488; DB 2; Length 666;  
Best Local Similarity 54.2%; Pred. No. 1.1e-21;  
Matches 84; Conservative 30; Mismatches 41; Indels 0; Gaps 0  
  
Qy 315 DEDISARFEDNEELRYSLRSTIERHAPVNRNIFIVTNGQIPSWLNDNPRVTIVTHQDVF 374  
Db DDKYDPSKFDKNELYRSLRSLKHAAMIRHYIVTNGQIPSWLDSYERYTVVPHEVLA 145  
  
Qy 375 RNLSHLPTFFSPALESHHRIEGLSQKFIYNDVWFQKDWPDFFYHSKGQKYYLTWP 434  
Db 146 PDPQLPTFFSSAJETFLHRIPKLSKRELYLNDIFLGAPLYPEDLYTEAGSVRYQAM 205  
  
Qy 435 VPNCAEGCPGSMKDGVCYCDKACNNSACDWDGDCS 469  
Db 206 VPGCALDCPWTYIGGACDRHCNIDACQFDGDCS 240  
  
RESULT 10  
Q8SX14 PRELIMINARY; PRT; 666 AA.  
ID AC Q8SX14;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE RE35033p.  
GN ORFNames=CG8027;  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,  
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,  
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,  
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
RA Celiniker S.;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY069618; AAU90356.1; -;  
DR FLYBASE; FBgn0033392; CG8027.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0030154; P:cell differentiation; IEA.  
DR InterPro; IPR000800; Notch\_region.  
DR Pfam; PF00066; Notch; 1.

[illegible]



```
Db 490 NPRERLAVDAEQTR-VKRLAGAYEG---KAYVAELLKERTHAPGVLLAERLEAVEVA 545
Qy 229 FLSGFPPTFKETNQLK-----TKLPENLSK-----VKLLQLYSE 263
Db 546 GLRVFRVWVSTRTRLRFEPAYGCDIEFRQVPEEGGQGFVAPLRPSAVGPKPLSLTPD 605
Qy 264 ASVALLKLNPKDFOELNKOTKK-----NMTIDGKELTISPAYLLND-----LSAI 309
Db 606 ARTFV-----KDRVPTLEPTKLVSDITFPDVAVTWVDDSDPRWQERRARRAALGLE 661
Qy 310 SQSKQDBDISASFEDNEELRYSLRSIERHAPWVRNIFIVTNGQIFSWNLNDNPRVTIYT 369
Db 662 AESSGDE--AARFRNRDELRYSLRSLAMPAPWIRKIYLVTDQTPFWLNTHEGIEVVS 718
Qy 370 HQDVERNLHLPTFSSPAIESHRIEGLSQEYLYLNDVWFGKDVWPDFFYSHSKGQV 429
Db 719 HRDIFTQDCDCLPTFNHSHIESQLHHDIGSEQEFLYLNDDVFICRPVGAORFFLPNGASRF 778
Qy 430 YLTWPVPCNAEGCPGSGWIKDGYCKACNNSA 460
Db 779 F--WSPTTVFVGEP-TEEDSGYFAAAKNRA 806
```

## RESULT 15

```
O69853 PRELIMINARY; PET; 586 AA.
AC O69853;
DT 01-AUG-1998 (T-EMBLrel. 07, Created)
DT 01-AUG-1998 (T-EMBLrel. 07, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Hypothetical protein SCO6023.
GN OrderedLocusNames=SCO6023; ORFNames=SCIC3.11;
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953; DOI=10.1038/417141a;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L.D., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K.M., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J.R., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL939126; CAA19235.1; -.
DR PIR; T34703; T34703.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 586 AA; 66009 MW; 75F98F0916484116 CRC64;
```

```
Query Match
Best Local Similarity 46.3%; Pred. No. 1.9e-11;
Matches 62; Conservative 30; Mismatches 33; Indels 9; Gaps 3;

Qy 290 IDGKELTISPAYLLW--DLSAISQSKQDEDISASREFNEELRYSLRSIERHAPWVRNIF 347
Db 282 VQND-----PA---WRRRSAYGGYHAGSANAARYISRLYSURALEQNAQAPWVRVH 334
Qy 348 IVTNGQIPSWNLNDNPRVTIVTHQDVFRLNLSHLPTFSSPAIESHRIEGLSKQFIYLN 407
Db 335 LVTDGQRPAWLNDSHRLTVVDHSEIFADPAALPTFNSHAIESRLHHIKGLSEHFIYLN 394
Qy 408 DVWFGKDVWDDDFY 421
Db 395 DMFLGRPVTPQDF 408
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2004, 12:38:18 ; Search time 37,6766 Seconds  
(without alignments)  
3122.980 Million cell updates/sec

Title: US-10-023-888-5

Perfect score: 1731

Sequence: 1 DTPADSLRYVKNILSKFGF.....RKIFPRRIKESPNIRIV 328

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_23Sep04:\*

1: Geneseq1980s:\*

2: Geneseq1990s:\*

3: Geneseq2000s:\*

4: Geneseq2001s:\*

5: Geneseq2002s:\*

6: Geneseq2003as:\*

7: Geneseq2003bs:\*

8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1731	100.0	328	4	ABR61377 Human Glc
2	1731	100.0	328	7	ADD27813 Human Glc
3	1731	100.0	328	7	ABW01489 Human Glc
4	1731	100.0	328	7	ABW01538 Human Glc
5	1731	100.0	1196	5	AAE25294 Human nuc
6	1731	100.0	1256	5	AAE25290 Human nuc
7	1698	98.1	328	4	ABR61380 Murine Glc
8	1698	98.1	328	7	ADD27818 GlcNAc-ph
9	1698	98.1	328	7	ABW01492 Mouse pro
10	1698	98.1	328	7	ABW01541 Mouse pro
11	1487	85.9	1199	7	ADD27810 Soluble h
12	1487	85.9	1199	7	ABW01487 N-acetylgl
13	1487	85.9	1199	7	ABW01536 N-acetylgl
14	1276	73.7	248	7	ADG39858 Protein s
15	1140	65.9	663	7	ADG39855 Protein s
16	1140	65.9	1459	6	ABU07381 Human pro
17	1140	65.9	1459	7	ADG39794 Human pro
18	704	40.7	132	4	AAE93369 Human pro
19	581	33.6	384	7	ADG39857 Protein s
20	506	29.2	652	4	ABE59094 Drosophil
21	495.5	28.6	651	7	ADG39859 Protein s
22	402	23.2	502	4	ABR61383 Partial D
23	402	23.2	502	7	ADD27824 Fruit fly
24	402	23.2	502	7	ABW01495 Partial f
25	402	23.2	502	7	ABW01544 Partial f

26	300	17.3	55	8	ABO57141 Human gen
27	145	8.4	28	4	ABR61390 Bovine Gl
28	145	8.4	28	4	AAE68261 Amino ter
29	124	7.2	77	4	AAE88395 Human imm
30	124	7.2	1312	5	ABW77984 Amino aci
31	124	7.2	1312	6	ABR53887 Protein s
32	124	7.2	1312	7	ADK63602 Disease c
33	113.5	6.6	545	2	AAW79296 Neisseria
34	112	6.5	46	7	ADD27834 GlcNAc-ph
35	106	6.1	574	5	ABP26069 Streptoco
36	101.5	5.9	3033	5	ABE49330 Listeria
37	101	5.8	2633	4	ABQ06505 Novel hum
38	101	5.8	2663	8	ADQ17932 Human sof
39	101	5.8	2688	4	AAE40883 Human poi
40	100.5	5.8	259	2	AAE79314 IL-3 cont
41	100.5	5.8	259	2	AAE79331 IL-3 cont
42	100.5	5.8	259	3	AAE53212 Human int
43	100.5	5.8	259	3	AAE53195 Human int
44	100.5	5.8	259	4	AAE14006 Myeloipoie
45	100.5	5.8	259	4	AAE13988 Myeloipoie

## ALIGNMENTS

## RESULT 1

ABR61377

ID ABR61377 standard; protein; 328 AA.

XX AC ABR61377;

XX DT 01-AUG-2003 (first entry)

XX DE Human GlcNAc-phosphotransferase beta-subunit.

XX KW Human; N-acetylglucosamine-1-phosphotransferase; nephrotropic;

XX KW GlcNAc-phosphotransferase; phosphodiester alpha-GlcNAcase;

XX KW N-acetylglucosamine-1-phosphodiester alpha-N-Acetylglucosaminidase;

XX KW enzyme replacement therapy; phosphorylated lysosomal hydrolase;

XX KW lysosomal storage disease; enzyme; beta-subunit.

XX OS Homo sapiens.

XX PN US6537785-B1.

XX PD 25-MAR-2003.

XX PF 10-AUG-2000; 2000US-00636077.

XX PR 14-SEP-1999; 99US-0153831P.

XX FA (GENZ-) GENZYME GLYCOBIOLOGY RES INST INC.

XX PI Canfield WM;

XX DR WPI; 2001-290356/30.

XX DR N-PSDB; ACCB1001.

XX PT Novel N-acetylglucosamine-1-phosphotransferase and N-acetylglucosamine-1-

XX PT phosphodiester alpha-N-Acetylglucosaminidase, useful for producing

XX PT phosphorylated lysosomal hydrolase for treating lysosomal storage

XX PS Disclosure; Page 28-29; 62pp; English.

XX CC The invention relates to a novel isolated human N-acetylglucosamine-1-

XX CC phosphotransferase (GlcNAc-phosphotransferase) (I) and phosphodiester

XX CC alpha-GlcNAcase (N-acetylglucosamine-1-phosphodiester alpha-N-

XX CC Acetylglucosaminidase) (II). The protein of the invention has

XX CC nephrotropic activity, and may be useful in enzyme replacement therapy. A

XX CC protein of the invention (I), (II) is useful for preparing a

XX CC phosphorylated lysosomal hydrolase. The phosphorylated hydrolase

XX CC comprising a terminal mannose-6-phosphate, is useful for treating a

CC patient suffering from a lysosomal storage disease. The present sequence  
 CC is used in the exemplification of the invention  
 XX  
 SQ Sequence 328 AA;  
 Query Match 100.0%; Score 1731; DB 4; Length 328;  
 Best Local Similarity 100.0%; Pred. No. 7,3e-160; Indels 0; Gaps 0;  
 Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DTFADSLRVNKLNSKFGFTSRKVPAPHPMIDRIVMQELQDMFPPEFDKTSFHKVRS 60  
 DB 1 DTFADSLRVNKLNSKFGFTSRKVPAPHPMIDRIVMQELQDMFPPEFDKTSFHKVRS 60  
 QY 61 EDMQFAFSFYIYMSAVQPLNISQVDFEVDTDQSGVLSDRIRTLATRIHELPLSLQDLT 120  
 DB 61 EDMQFAFSFYIYMSAVQPLNISQVDFEVDTDQSGVLSDRIRTLATRIHELPLSLQDLT 120  
 QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDNPPLPVTKSLVNTCKEPTDKIKHAYKDK 180  
 DB 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDNPPLPVTKSLVNTCKEPTDKIKHAYKDK 180  
 QY 181 NKYRFEIMGEESIAFKMIRTNVSHVVGQDDIRKNPRKFCVCLNDNDHNHKAQOTVKAVL 240  
 DB 181 NKYRFEIMGEESIAFKMIRTNVSHVVGQDDIRKNPRKFCVCLNDNDHNHKAQOTVKAVL 240  
 QY 241 RQFYSMFPIPSQFELPREYRNRFLHMHLEQWRAVYRDKLKFVTHCVLATLIMFTIFSFF 300  
 DB 241 RQFYSMFPIPSQFELPREYRNRFLHMHLEQWRAVYRDKLKFVTHCVLATLIMFTIFSFF 300  
 QY 301 AEQIALKKEKIFPRRIHKEASPNRIRV 328  
 DB 301 AEQIALKKEKIFPRRIHKEASPNRIRV 328  
 RESULT 3  
 ABW01489  
 ID ABW01489 standard; protein; 328 AA.  
 AC ABW01489;  
 XX  
 DT 15-JAN-2004 (first entry)  
 DE Human GlcNAc-phosphotransferase beta subunit precursor protein.  
 DE Mannose glycoprotein; gene therapy; carbohydrate deficient cell;  
 KW lysosomal storage disease; N-acetylglucosamine-1-phosphotransferase;  
 KW gastrointestinal; human; enzyme; lectin resistant cell;  
 KW deoxymannojirimycin; kifunensine; glycosylation inhibition.  
 OS Homo sapiens.  
 XX US2003124652-A1.  
 XX  
 PD 03-JUL-2003.  
 XX  
 PF 21-DEC-2001; 2001US-00023889.  
 XX  
 PR 21-DEC-2001; 2001US-00023889.  
 XX  
 PA (NOVA-) NOVAZYME PHARM INC.  
 XX  
 PI Canfield WM;  
 XX  
 XX WPI; 2003-810984/76.  
 DR N-PSDB; RAD62491.  
 XX  
 PT Producing a high mannose glycoprotein for treating lysosomal storage  
 PT disease, comprises culturing the lectin resistant mammalian cell in the  
 PT presence of deoxymannojirimycin and kifunensine.  
 XX  
 PS Claim 10; Page 18-19; 46pp; English.

CC patient suffering from a lysosomal storage disease. The present sequence  
 CC is used in the exemplification of the invention  
 XX  
 SQ Sequence 328 AA;  
 Query Match 100.0%; Score 1731; DB 4; Length 328;  
 Best Local Similarity 100.0%; Pred. No. 7,3e-160; Indels 0; Gaps 0;  
 Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DTFADSLRVNKLNSKFGFTSRKVPAPHPMIDRIVMQELQDMFPPEFDKTSFHKVRS 60  
 DB 1 DTFADSLRVNKLNSKFGFTSRKVPAPHPMIDRIVMQELQDMFPPEFDKTSFHKVRS 60  
 QY 61 EDMQFAFSFYIYMSAVQPLNISQVDFEVDTDQSGVLSDRIRTLATRIHELPLSLQDLT 120  
 DB 61 EDMQFAFSFYIYMSAVQPLNISQVDFEVDTDQSGVLSDRIRTLATRIHELPLSLQDLT 120  
 QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDNPPLPVTKSLVNTCKEPTDKIKHAYKDK 180  
 DB 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDNPPLPVTKSLVNTCKEPTDKIKHAYKDK 180  
 QY 181 NKYRFEIMGEESIAFKMIRTNVSHVVGQDDIRKNPRKFCVCLNDNDHNHKAQOTVKAVL 240  
 DB 181 NKYRFEIMGEESIAFKMIRTNVSHVVGQDDIRKNPRKFCVCLNDNDHNHKAQOTVKAVL 240  
 QY 241 RQFYSMFPIPSQFELPREYRNRFLHMHLEQWRAVYRDKLKFVTHCVLATLIMFTIFSFF 300  
 DB 241 RQFYSMFPIPSQFELPREYRNRFLHMHLEQWRAVYRDKLKFVTHCVLATLIMFTIFSFF 300  
 QY 301 AEQIALKKEKIFPRRIHKEASPNRIRV 328  
 DB 301 AEQIALKKEKIFPRRIHKEASPNRIRV 328  
 RESULT 2  
 ADD27813  
 ID ADD27813 standard; protein; 328 AA.  
 AC ADD27813;  
 XX  
 DT 15-JAN-2004 (first entry)  
 DE Human GlcNAc-phosphotransferase beta subunit.  
 DE human; protein phosphorylation; soluble GlcNAc-phosphotransferase;  
 KW UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease.  
 OS Homo sapiens.  
 XX US2003119088-A1.  
 XX  
 PD 26-JUN-2003.  
 XX  
 PF 21-DEC-2001; 2001US-00023888.  
 XX  
 PR 21-DEC-2001; 2001US-00023888.  
 XX  
 PA (NOVA-) NOVAZYME PHARM INC.  
 XX  
 PI Canfield W, Kudo M;  
 XX  
 XX WPI; 2003-801323/75.  
 DR N-PSDB; ADD27811.  
 XX  
 PT Phosphorylating a protein for treating a patient suffering from a  
 PT lysosomal storage disease e.g. Fabry's disease by contacting the protein  
 PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated  
 PT protein.  
 XX  
 PS Claim 8; SEQ ID NO 5; 55pp; English.  
 CC  
 CC The invention relates to a method of phosphorylating a protein comprising  
 CC contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-



XX The invention relates to a method for producing a high mannose  
 CC glycoprotein. The method comprises: introducing and expressing a  
 CC polynucleotide encoding a glycoprotein into a mammalian cell; culturing  
 CC the cell in the presence of a lectin to obtain a lectin resistant cell;  
 CC isolating the cell; culturing the cell in the presence of  
 CC deoxymannojirimycin and kifunensine to inhibit glycosylation of the  
 CC glycoprotein; and collecting the glycoprotein. The invention is useful in  
 CC gene therapy. The method is useful for producing a high mannose  
 CC glycoprotein in a complex carbohydrate deficient cell for treating  
 CC lysosomal storage disease. The present sequence is human N-  
 CC acetylglucosamine-1 (GlcNAc)-phosphotransferase beta subunit precursor  
 CC protein  
 XX  
 SQ Sequence 328 AA;

Query Match 100.0%; Score 1731; DB 7; Length 328;  
 Best Local Similarity 100.0%; Pred. No. 7.3e-160;  
 Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DTFADSLRYVVKILNSKFGFTSRKVPAPMHPMIDRIYVQELQDMFPEEDKTSFHKVRS 60  
 Db 1 DTFADSLRYVVKILNSKFGFTSRKVPAPMHPMIDRIYVQELQDMFPEEDKTSFHKVRS 60  
 QY 61 EDMQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDRERTLATRIHELPLSLQDLT 120  
 Db 61 EDMQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDRERTLATRIHELPLSLQDLT 120  
 QY 121 GLEHMLNCSKMLPADITQLNNIPPTQESYYDPLPPVTKSLVNTCKPVTDKIHAYKDK 180  
 Db 121 GLEHMLNCSKMLPADITQLNNIPPTQESYYDPLPPVTKSLVNTCKPVTDKIHAYKDK 180  
 QY 181 NKYFEIMGEBEIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDINHNHKAQTVKAVL 240  
 Db 181 NKYFEIMGEBEIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDINHNHKAQTVKAVL 240  
 QY 241 RDFSFMFPPIPSQFELPREYRNRFLEHMQEWRAYRDKLKFVTHCVLATLIMFTIFSFF 300  
 Db 241 RDFSFMFPPIPSQFELPREYRNRFLEHMQEWRAYRDKLKFVTHCVLATLIMFTIFSFF 300  
 QY 301 AEQIALKRRKIFPRRRIHKEASPNRIRV 328  
 Db 301 AEQIALKRRKIFPRRRIHKEASPNRIRV 328

RESULT 4  
 ABW01538  
 ID ABW01538 standard; protein; 328 AA.  
 AC ABW01538;

XX  
 XX  
 DT 15-JAN-2004 (first entry)  
 XX

DE Human GlcNAc-phosphotransferase beta subunit precursor protein.

XX Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;  
 KW N-acetylglucosamine-1-phosphotransferase; enzyme; gene therapy; human.

XX Homo sapiens.

XX US2003124653-A1.

XX 03-JUL-2003.

XX 21-DEC-2001; 2001US-00023890.

XX 21-DEC-2001; 2001US-00023890.

XX (NOVA-) NOVAZYME PHARM INC.

XX Canfield WM;

XX WPI; 2003-810985/76.

DR N-PSDB; AAD62650.

XX Producing a glycoprotein with reduced complex carbohydrates by culturing  
 PT the lectin resistant mammalian cell expressing the glycoprotein for  
 PT treating lysosomal storage disease.

XX Claim 10; Page 18-19; 46pp; English.

XX The present invention provides a method of producing a glycoprotein  
 CC having reduced complex carbohydrates by culturing the lectin resistant  
 CC mammalian cell expressing the glycoprotein. The method is useful for  
 CC producing a glycoprotein with reduced complex carbohydrates for treating  
 CC lysosomal storage disease. The present invention is also useful in gene  
 CC therapy. The present sequence is human N-acetylglucosamine-1 (GlcNAc)-  
 CC phosphotransferase beta subunit precursor protein  
 XX

SQ Sequence 328 AA;

Query Match 100.0%; Score 1731; DB 7; Length 328;  
 Best Local Similarity 100.0%; Pred. No. 7.3e-160;  
 Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVVKILNSKFGFTSRKVPAPMHPMIDRIYVQELQDMFPEEDKTSFHKVRS 60  
 Db 1 DTFADSLRYVVKILNSKFGFTSRKVPAPMHPMIDRIYVQELQDMFPEEDKTSFHKVRS 60

QY 61 EDMQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDRERTLATRIHELPLSLQDLT 120  
 Db 61 EDMQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDRERTLATRIHELPLSLQDLT 120

QY 121 GLEHMLNCSKMLPADITQLNNIPPTQESYYDPLPPVTKSLVNTCKPVTDKIHAYKDK 180  
 Db 121 GLEHMLNCSKMLPADITQLNNIPPTQESYYDPLPPVTKSLVNTCKPVTDKIHAYKDK 180

QY 181 NKYFEIMGEBEIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDINHNHKAQTVKAVL 240  
 Db 181 NKYFEIMGEBEIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDINHNHKAQTVKAVL 240

QY 241 RDFSFMFPPIPSQFELPREYRNRFLEHMQEWRAYRDKLKFVTHCVLATLIMFTIFSFF 300  
 Db 241 RDFSFMFPPIPSQFELPREYRNRFLEHMQEWRAYRDKLKFVTHCVLATLIMFTIFSFF 300

QY 301 AEQIALKRRKIFPRRRIHKEASPNRIRV 328

Db 301 AEQIALKRRKIFPRRRIHKEASPNRIRV 328

RESULT 5

AAE25294

ID AAE25294 standard; protein; 1196 AA.

XX  
 AC AAE25294;

XX  
 DT 30-OCT-2002 (first entry)

XX Human nucleic acid-associated protein (NAAP-13).

XX Human; nucleic acid-associated protein; NAAP-13; neurological disorder;  
 KW arteriosclerosis; cancer; cell proliferative disorder; atherosclerosis;  
 KW lymphoma; epilepsy; Alzheimer's disease; developmental; anticonvulsant;  
 KW autoimmune disorder; AIDS; allergy; anaemia; stroke; malaria; leishmania;  
 KW gene therapy; neurotropic; neuroprotective; cerebroprotective; virucide;  
 KW immunosuppressive; protozoaside; antimicrobial.

XX Homo sapiens.

XX Key Location/Qualifiers

XX 1..34 /label= Signal\_peptide

XX 1..19 /note= "Cytosolic domain"

XX 20..42 /note= "Transmembrane domain"

XX  
 FT

FT Protein 35. .1196 /note= "Mature human NAAP-13"

FT Domain 43. .1152 /note= "Non-cytosolic domain"

FT Domain 1153. .1175 /note= "Transmembrane domain"

FT Domain 1176. .1196 /note= "Cytosolic domain"

XX WO200250279-A2.

XX 27-JUN-2002.

XX 19-DEC-2001; 2001WO-US050256.

XX 21-DEC-2000; 2000US-0257714P.

XX 05-JAN-2001; 2001US-0260081P.

XX 16-JAN-2001; 2001US-0262302P.

XX 23-JAN-2001; 2001US-0263823P.

XX 02-FEB-2001; 2001US-0266088P.

XX 29-OCT-2001; 2001US-0348442P.

XX (INCY-) INCYTE GENOMICS INC.

XX Baughin MR, Lu Y, Arvizu C, Ramkumar J, Yao MG, Policky JL;

XX Wallia NK, Tribouley KM, Yue H, Batra S, Ding L, Lal PG;

XX Borowsky ML, Lu DAM, Gandhi AR, Griffin JA, Xu Y, Azimzai Y;

XX Gietzen KJ, Tang YT, Warren BA, Mason PM, Hatfield AJA;

XX Lee EA, Yang J, Gorvad AE, Emerling BM, Marquis JP, Lee SY;

XX Swarnakar A, Reddy R;

XX WPI; 2002-519887/55.

XX N-PSDB; AAD41203.

XX Nucleic acid associated proteins and nucleic acids for diagnosing,

XX treating and preventing cell proliferative (e.g. cancers), neurological

XX (e.g. epilepsy or stroke) or autoimmune disorders (e.g. AIDS).

XX Claim 68; Page 169-172; 193pp; English.

XX The invention relates to nucleic acid-associated proteins (NAAP) and

XX nucleic acids. The nucleic acid and amino acid sequences are useful for

XX diagnosing, treating and preventing cell proliferative e.g.

XX arteriosclerosis, atherosclerosis, lymphoma or cancers, neurological

XX (e.g. epilepsy, Alzheimer's disease or stroke), developmental, and

XX autoimmune disorders (e.g. AIDS, allergies, or anaemia) or infections

XX (e.g. malaria, or leishmania), as well as in assessing the effects of

XX exogenous compound on the expression of nucleic acid and amino acid

XX sequences of nucleic acid-associated proteins. The invention is useful in

XX gene therapy. The present sequence is human NAAP-13

XX Sequence 1196 AA;

XX Query Match 100.0%; Score 1731; DB 5; Length 1196;

XX Best Local Similarity 100.0%; Pred. No. 4.8e-159;

XX Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYNKNILNSKFGFTSRKVPAPMPHMDIRIVMQLQDMFPPEPKTSFHKVRHS 60

DB 869 DTFADSLRYNKNILNSKFGFTSRKVPAPMPHMDIRIVMQLQDMFPPEPKTSFHKVRHS 928

QY 61 EDMQFASFYFYLMSAVQPLNISQVDEVDTDQSGVLSDRIRTLARIRHELPLSLQDLT 120

DB 929 EDMQFASFYFYLMSAVQPLNISQVDEVDTDQSGVLSDRIRTLARIRHELPLSLQDLT 988

QY 121 GLEHMLNCSKMLPADITQLNNIPPTQESYDPLNLPVTKSLVTNCKPVTDXIHKAYKDK 180

DB 989 GLEHMLNCSKMLPADITQLNNIPPTQESYDPLNLPVTKSLVTNCKPVTDXIHKAYKDK 1048

QY 191 NKYRFEIMGEIEIAFKMRTNVSHVVGOLDIRKNPKRFVCLNDNIDHNHKAQTVKAVL 240

DB 1049 NKYRFEIMGEIEIAFKMRTNVSHVVGOLDIRKNPKRFVCLNDNIDHNHKAQTVKAVL 1108

QY 241 RDFYESMFPPIPSQFELPREYRNRLFMHHELOEWRAYRDKLKFTHCVLATLIMFTISFF 300

DB 1109 RDFYESMFPPIPSQFELPREYRNRLFMHHELOEWRAYRDKLKFTHCVLATLIMFTISFF 1168

QY 301 AEQIALKRKIFPRRRIRHKEASPNRIRV 328

DB 1169 AEQIALKRKIFPRRRIRHKEASPNRIRV 1196

RESULT 6

AAE25290

ID AAE25290 standard; protein; 1256 AA.

XX

XX AAE25290;

XX 30-OCT-2002 (first entry)

XX Human nucleic acid-associated protein (NAAP-9).

XX Human; nucleic acid-associated protein; NAAP-9; neurological disorder;

XX arteriosclerosis; cancer; cell proliferative disorder; atherosclerosis;

XX lymphoma; epilepsy; Alzheimer's disease; developmental; anticonvulsant;

XX autoimmune disorder; AIDS; allergy; anaemia; stroke; malaria; leishmania;

XX gene therapy; nontropic; neuroprotective; cerebroprotective; virucide;

XX immunosuppressive; protozoacide; antimicrobial.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Domain 448. .469 /note= "Notch domain"

XX Domain 500. .536 /note= "Notch domain"

XX Domain 1018. .1030 /note= "EF-hand calcium-binding domain"

XX WO200250279-A2.

XX 27-JUN-2002.

XX 19-DEC-2001; 2001WO-US050256.

XX 21-DEC-2000; 2000US-0257714P.

XX 05-JAN-2001; 2001US-0260081P.

XX 16-JAN-2001; 2001US-0262302P.

XX 23-JAN-2001; 2001US-0263823P.

XX 02-FEB-2001; 2001US-0266088P.

XX 29-OCT-2001; 2001US-0348442P.

XX (INCY-) INCYTE GENOMICS INC.

XX Baughin MR, Lu Y, Arvizu C, Ramkumar J, Yao MG, Policky JL;

XX Wallia NK, Tribouley KM, Yue H, Batra S, Ding L, Lal PG;

XX Borowsky ML, Lu DAM, Gandhi AR, Griffin JA, Xu Y, Azimzai Y;

XX Gietzen KJ, Tang YT, Warren BA, Mason PM, Hatfield AJA;

XX Lee EA, Yang J, Gorvad AE, Emerling BM, Marquis JP, Lee SY;

XX Swarnakar A, Reddy R;

XX WPI; 2002-519887/55.

XX N-PSDB; AAD41199.

XX Nucleic acid associated proteins and nucleic acids for diagnosing,

XX treating and preventing cell proliferative (e.g. cancers), neurological

XX (e.g. epilepsy or stroke) or autoimmune disorders (e.g. AIDS).

XX Claim 64; Page 163-165; 193pp; English.

XX The invention relates to nucleic acid-associated proteins (NAAP) and

XX nucleic acids. The nucleic acid and amino acid sequences are useful for

XX diagnosing, treating and preventing cell proliferative e.g.

XX arteriosclerosis, atherosclerosis, lymphoma or cancers, neurological

XX (e.g. epilepsy, Alzheimer's disease or stroke), developmental, and

XX autoimmune disorders (e.g. AIDS, allergies, or anaemia) or infections



XX Phosphorylating a protein for treating a patient suffering from a  
 PT lysosomal storage disease e.g. Fabry's disease by contacting the protein  
 PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated  
 PT protein.  
 XX  
 XX Disclosure; SEQ ID NO 10; 55pp; English.  
 XX  
 XX The invention relates to a method of phosphorylating a protein comprising  
 CC contacting the protein with a soluble GlcNAc-phosphotransferase (GDP-N-  
 CC acetylglucosamine) and producing a phosphorylated protein. The method is  
 CC useful for treating a patient suffering from a lysosomal storage disease  
 CC e.g. Fabry's disease. The present sequence represents the amino acid  
 CC sequence of a GlcNAc-phosphotransferase associated protein.  
 XX  
 XX Sequence 328 AA;

Query Match 98.1%; Score 1698; DB 7; Length 328;  
 Best Local Similarity 97.6%; Pred. No. 1.2e-156;  
 Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPMPHMDIRIVMQLQDMFPEEDKTSFHKVRS 60  
 DB |||||  
 QY 61 EDMQFAFSFYLLMSAVQPLNISQVFDEVDTPDQSGVLSDEIRTLATRIHELPLSLQDLT 120  
 DB |||||  
 QY 61 EDMQFAFSFYLLMSAVQPLNISQVFDEVDTPDQSGVLSDEIRTLATRIHELPLSLQDLT 120  
 DB |||||  
 QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYVDNLPVTKSLVTKNCPVTDKHKAYKDK 180  
 DB |||||  
 QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYVDNLPVTKSLVTKNCPVTDKHKAYKDK 180  
 DB |||||  
 QY 181 NKYRFEIMGEEIEAFKMRITNVSHVVGQDDIRKNPKRFVCLNDNDHNHDKAQTVKAVL 240  
 DB |||||  
 QY 181 NKYRFEIMGEEIEAFKMRITNVSHVVGQDDIRKNPKRFVCLNDNDHNHDKAQTVKAVL 240  
 DB |||||  
 QY 241 RDFYESMPPIPSQFELPREYRNRFHMHLEQWRAVDKLFKFWTHCVLATLIMFTTISFF 300  
 DB |||||  
 QY 241 RDFYESMPPIPSQFELPREYRNRFHMHLEQWRAVDKLFKFWTHCVLATLIMFTTISFF 300  
 DB |||||  
 QY 301 AEQIALKRRKIFPRRRIHKEASPNRIRV 328  
 DB |||||  
 QY 301 AEQIALKRRKIFPRRRIHKEASPNRIRV 328  
 DB |||||

RESULT 9  
 ABW01492  
 ID ABW01492 standard; protein; 328 AA.  
 AC ABW01492;  
 XX

15-JAN-2004 (first entry)

Mouse protein #2 used to illustrate the method of the invention.  
 Mannose glycoprotein; gene therapy: carbohydrate deficient cell;  
 lysosomal storage disease; gastrointestinal; mouse;  
 lectin resistant cell; deoxymannojirimycin; kifunensine;  
 glycosylation inhibition.

Mus musculus.

US2003124652-A1.

03-JUL-2003.

21-DEC-2001; 2001US-00023889.

21-DEC-2001; 2001US-00023889.

(NOVA-) NOVAZIME PHARM INC.

PI Canfield WM;  
 XX  
 XX WPI; 2003-810984/76.  
 XX  
 PT Producing a high mannose glycoprotein for treating lysosomal storage  
 PT disease, comprises culturing the lectin resistant mammalian cell in the  
 PT presence of deoxymannojirimycin and kifunensine.  
 XX  
 XX Disclosure; Page 25-26; 46pp; English.  
 XX  
 XX The invention relates to a method for producing a high mannose  
 CC glycoprotein. The method comprises: introducing and expressing a  
 CC polynucleotide encoding a glycoprotein into a mammalian cell; culturing  
 CC the cell in the presence of a lectin to obtain a lectin resistant cell;  
 CC isolating the cell; culturing the cell in the presence of  
 CC deoxymannojirimycin and kifunensine to inhibit glycosylation of the  
 CC glycoprotein; and collecting the glycoprotein. The invention is useful in  
 CC gene therapy. The method is useful for producing a high mannose  
 CC glycoprotein in a complex carbohydrate deficient cell for treating  
 CC lysosomal storage disease. The present sequence is mouse protein used to  
 CC illustrate the method of the invention  
 XX  
 XX Sequence 328 AA;

Query Match 98.1%; Score 1698; DB 7; Length 328;  
 Best Local Similarity 97.6%; Pred. No. 1.2e-156;  
 Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPMPHMDIRIVMQLQDMFPEEDKTSFHKVRS 60  
 DB |||||  
 QY 61 EDMQFAFSFYLLMSAVQPLNISQVFDEVDTPDQSGVLSDEIRTLATRIHELPLSLQDLT 120  
 DB |||||  
 QY 61 EDMQFAFSFYLLMSAVQPLNISQVFDEVDTPDQSGVLSDEIRTLATRIHELPLSLQDLT 120  
 DB |||||  
 QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYVDNLPVTKSLVTKNCPVTDKHKAYKDK 180  
 DB |||||  
 QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYVDNLPVTKSLVTKNCPVTDKHKAYKDK 180  
 DB |||||  
 QY 181 NKYRFEIMGEEIEAFKMRITNVSHVVGQDDIRKNPKRFVCLNDNDHNHDKAQTVKAVL 240  
 DB |||||  
 QY 181 NKYRFEIMGEEIEAFKMRITNVSHVVGQDDIRKNPKRFVCLNDNDHNHDKAQTVKAVL 240  
 DB |||||  
 QY 241 RDFYESMPPIPSQFELPREYRNRFHMHLEQWRAVDKLFKFWTHCVLATLIMFTTISFF 300  
 DB |||||  
 QY 241 RDFYESMPPIPSQFELPREYRNRFHMHLEQWRAVDKLFKFWTHCVLATLIMFTTISFF 300  
 DB |||||  
 QY 301 AEQIALKRRKIFPRRRIHKEASPNRIRV 328  
 DB |||||  
 QY 301 AEQIALKRRKIFPRRRIHKEASPNRIRV 328  
 DB |||||

RESULT 10  
 ABW01541  
 ID ABW01541 standard; protein; 328 AA.  
 XX  
 AC ABW01541;  
 XX

15-JAN-2004 (first entry)

Mouse protein #2 used to illustrate the method of the invention.

Glycoprotein; lectin; lysosomal storage disease; gastrointestinal;  
 N-acetylglucosamine-1-phosphotransferase; gene therapy; mouse.

Mus musculus.

US2003124653-A1.

03-JUL-2003.

21-DEC-2001; 2001US-00023890.

XX PR 21-DEC-2001; 2001US-00023890.  
XX PA (NOVA-) NOVAZYME PHARM INC.  
XX PI Canfield W;  
XX XX WPI; 2003-810985/76.  
XX PT Producing a glycoprotein with reduced complex carbohydrates by culturing  
PT the lectin resistant mammalian cell expressing the glycoprotein for  
PT treating lysosomal storage disease.  
XX PS Disclosure; Page 25-26; 46pp; English.  
XX CC The present invention provides a method of producing a glycoprotein  
CC having reduced complex carbohydrates by culturing the lectin resistant  
CC mammalian cell expressing the glycoprotein. The method is useful for  
CC producing a glycoprotein with reduced complex carbohydrates for treating  
CC lysosomal storage disease. The present invention is also useful in gene  
CC therapy. The present sequence is mouse protein used to illustrate the  
CC method of the invention  
XX SQ Sequence 328 AA;  
Query Match 98.1%; Score 1698; DB 7; Length 328;  
Best Local Similarity 97.6%; Pred. No. 1.2e-156;  
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;  
QY 1 DTFADSLRYVNKILNSKFGFTSRKVPAPMHPMIDRIVMQELQDMPEEFDKTSFHKVRS 60  
Db 1 DTFADSLRYVNKILNSKFGFTSRKVPAPMHPMIDRIVMQELQDMPEEFDKTSFHKVRS 60  
QY 61 EDMQFAFSFYFLMSAVQPLNISQVFEVDTDQSGVLSDRERTLATRIHELPLSLQDLT 120  
Db 61 EDMQFAFSFYFLMSAVQPLNISQVFEVDTDQSGVLSDRERTLATRIHELPLSLQDLT 120  
QY 121 GLEHMLNCSKMLPADITQLANNIPPTQESYYDPNLPVTKSLVNTCKPVTDKHKAYKDK 180  
Db 121 GLEHMLNCSKMLPANITQLANNIPPTQEAAYDPNLPVTKSLVNTCKPVTDKHKAYKDK 180  
QY 181 NKYFEIMGEBEIAFKMIRTNVSHVVGQDDIRKNPKRFVCLNDNDIDHNHKAQTVKAVL 240  
Db 181 NKYFEIMGEBEIAFKMIRTNVSHVVGQDDIRKNPKRFVCLNDNDIDHNHKAQTVKAVL 240  
QY 241 RDFSFMFPPIPSQFELPREYRNRFHMHLEQWRAEDKLFKFWTHCVLATLIIFTIFSFF 300  
Db 241 RDFSFMFPPIPSQFELPREYRNRFHMHLEQWRAEDKLFKFWTHCVLATLIIFTIFSFF 300  
QY 301 AEQIILKRRKIFPRRRHKASPRIRV 328  
Db 301 AEQIILKRRKIFPRRRHKASPRIRV 328  
RESULT 11  
ADD27810  
ID ADD27810 standard; protein; 1199 AA.  
AC ADD27810;  
XX XX  
DT 15-JAN-2004 (first entry)  
XX XX  
DE Soluble human GlcNAc-phosphotransferase.  
XX human; protein phosphorylation; soluble GlcNAc-phosphotransferase;  
KW UDP-N-acetylglucosamine; lysosomal storage disease; Fabry's disease;  
KW enzyme.  
XX Synthetic.  
OS Homo sapiens.  
XX US2003119088-A1.  
XX

PD 26-JUN-2003.  
XX 21-DEC-2001; 2001US-00023888.  
XX 21-DEC-2001; 2001US-00023888.  
XX (NOVA-) NOVAZYME PHARM INC.  
XX PA Canfield W, Kudo M;  
XX PI WPI; 2003-801323/75.  
XX DR N-PSDB; ADD27809.  
XX PT Phosphorylating a protein for treating a patient suffering from a  
PT lysosomal storage disease e.g. Fabry's disease by contacting the protein  
PT with a soluble GlcNAc-phosphotransferase and producing a phosphorylated  
PT protein.  
XX PS Claim 3; SEQ ID NO 2; 55pp; English.  
XX CC The invention relates to a method of phosphorylating a protein comprising  
CC contacting the protein with a soluble GlcNAc-phosphotransferase (UDP-N-  
CC acetylglucosamine) and producing a phosphorylated protein. The method is  
CC useful for treating a patient suffering from a lysosomal storage disease  
CC e.g. Fabry's disease. The present sequence represents the amino acid  
CC sequence of soluble human GlcNAc-phosphotransferase.  
XX SQ Sequence 1199 AA;  
Query Match 85.9%; Score 1487; DB 7; Length 1199;  
Best Local Similarity 100.0%; Pred. No. 2.9e-135;  
Matches 281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DTFADSLRYVNKILNSKFGFTSRKVPAPMHPMIDRIVMQELQDMPEEFDKTSFHKVRS 60  
Db 919 DTFADSLRYVNKILNSKFGFTSRKVPAPMHPMIDRIVMQELQDMPEEFDKTSFHKVRS 978  
QY 61 EDMQFAFSFYFLMSAVQPLNISQVFEVDTDQSGVLSDRERTLATRIHELPLSLQDLT 120  
Db 979 EDMQFAFSFYFLMSAVQPLNISQVFEVDTDQSGVLSDRERTLATRIHELPLSLQDLT 1038  
QY 121 GLEHMLNCSKMLPADITQLANNIPPTQESYYDPNLPVTKSLVNTCKPVTDKHKAYKDK 180  
Db 1039 GLEHMLNCSKMLPADITQLANNIPPTQESYYDPNLPVTKSLVNTCKPVTDKHKAYKDK 1098  
QY 181 NKYFEIMGEBEIAFKMIRTNVSHVVGQDDIRKNPKRFVCLNDNDIDHNHKAQTVKAVL 240  
Db 1099 NKYFEIMGEBEIAFKMIRTNVSHVVGQDDIRKNPKRFVCLNDNDIDHNHKAQTVKAVL 1158  
QY 241 RDFSFMFPPIPSQFELPREYRNRFHMHLEQWRAEDKLFK 281  
Db 1159 RDFSFMFPPIPSQFELPREYRNRFHMHLEQWRAEDKLFK 1199  
RESULT 12  
ABW01487  
ID ABW01487 standard; protein; 1199 AA.  
AC ABW01487;  
XX XX  
DT 15-JAN-2004 (first entry)  
XX XX  
DE N-acetylglucosamine-1 (GlcNAc)-phosphotransferase.  
XX Mannose glycoprotein; gene therapy; carbohydrate deficient cell;  
KW lysosomal storage disease; N-acetylglucosamine-1-phosphotransferase;  
KW gastrointestinal; enzyme; lectin resistant cell; deoxymannojirimycin;  
KW kifunensine; glycosylation inhibition.  
XX Unidentified.  
OS US2003124652-A1.  
XX





KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
 KW immune disorder; haematopoietic disorder; dyslipidaemia;  
 KW wasting disorder.  
 XX Homo sapiens.  
 XX US2003203843-A1.  
 XX 30-OCT-2003.  
 XX  
 XX 11-APR-2002; 2002US-00120801.  
 XX  
 XX 20-APR-2001; 2001US-0285609P.  
 XX 23-APR-2001; 2001US-0285748P.  
 XX 24-APR-2001; 2001US-0286068P.  
 XX 25-APR-2001; 2001US-0286292P.  
 XX 03-MAY-2001; 2001US-0288334P.  
 XX 16-MAY-2001; 2001US-0291241P.  
 XX 14-SEP-2001; 2001US-0322284P.  
 XX  
 XX (PENA/) PENA C E A.  
 XX (GUOX/) GUO X.  
 XX (SHIM/) SHIMKETS R A.  
 XX (PADI/) PADIGARU M.  
 XX (KEKU/) KEKUDA R.  
 XX (SPYT/) SPYTEK K A.  
 XX (MEHR/) MEHRABAN F.  
 XX (TOPP/) TOPPER J N.  
 XX (MALY/) MALYANKAR U M.  
 XX (WASS/) WASSEMAN S M.  
 XX (EDIN/) EDINGER S R.  
 XX (SMIT/) SMITHSON G.  
 XX (GUNT/) GUNTHER E.  
 XX (KOMU/) KOMUVES L.  
 XX  
 XX PENA CEA, Guo X, Shimkets RA, Padigar M, Kekuda R, Spytek KA;  
 XX Mehrahan F, Topper JN, Malyankar UM, Wasserman SM, Edinger SR,  
 XX Smithson G, Gunther E, Komuves L;  
 XX WPI: 2003-900671/82.  
 XX  
 XX New NOVX polypeptides and nucleic acids, useful for diagnosing or  
 XX treating e.g. cardiomyopathy, atherosclerosis, hypertension, scleroderma,  
 XX obesity, prostate cancer, AIDS, bronchial asthma, Crohn's disease, or  
 XX multiple sclerosis.  
 XX  
 XX Disclosure; SEQ ID NO 90; 215pp; English.  
 XX  
 XX The invention relates to a new isolated polypeptide comprising an amino  
 XX acid sequence selected from 17 fully defined human NOVX sequences (even  
 XX SEQ ID NOS between ADG39770 and ADG39802), a mature form of the NOVX  
 XX amino acid or a variant of NOVX, where one or more amino acid residue in  
 XX the variant differs in no more than 15% of the amino acid residues of  
 XX NOVX. Also included are an isolated nucleic acid (NA) molecule  
 XX (comprising a nucleic acid sequence encoding a NOVX polypeptide above  
 XX (odd SEQ ID NOS between ADG39769 and ADG39801), a nucleic acid fragment  
 XX encoding at least a portion of a NOVX polypeptide and a complement of  
 XX NOVX NA), a vector comprising NOVX NA, a cell comprising the vector, an  
 XX antibody that immunospecifically binds to NOVX, a method for determining  
 XX the presence or amount of NOVX or NOVX NA in a sample, a method of  
 XX identifying an agent that binds to NOVX, a method for identifying an  
 XX agent that modulates the expression or activity of NOVX, a method for  
 XX modulating the activity of NOVX, a method of treating or preventing a  
 XX NOVX-associated disorder, a method for screening for a modulator of  
 XX activity or of latency or predisposition to a NOVX-associated disorder, a  
 XX method for determining the presence of or predisposition to a disease  
 XX associated with altered levels of NOVX or NOVX NA in a first mammalian  
 XX subject and a method of treating a pathological state in a mammal by  
 XX administering NOVX or an antibody that binds to NOVX. The NOVX  
 XX polypeptide, nucleic acid or antibody is useful in the manufacture of a  
 XX medicament for treating a syndrome associated with a human disease or a  
 XX NOVX-associated disorder. The NOVX polypeptides and nucleic acids  
 XX encoding them are useful for diagnosing or treating pathologies, diseases

CC or conditions associated with NOVX sequences, including cardiomyopathy,  
 CC atherosclerosis, hypertension, congenital heart defects, pulmonary  
 CC stenosis, scleroderma, obesity, metabolic disturbances associated with  
 CC obesity, transplantation, adrenoleukodystrophy, congenital adrenal  
 CC hyperplasia, prostate cancer, diabetes, metabolic disorders, neoplasm,  
 CC adenocarcinoma, fertility, haemophilia, graft versus host disease, AIDS,  
 CC bronchial asthma, Crohn's disease, multiple sclerosis, infectious  
 CC disease, anorexia, neurodegenerative disorders (e.g. Alzheimer's disease,  
 CC or Parkinson's disease), immune disorders, haematopoietic disorders,  
 CC dyslipidaemias, and wasting disorders associated with chronic diseases.  
 CC The polypeptides can be used as immunogens to produce antibodies and as  
 CC vaccines. The sequences may further be used in chromosome mapping,  
 CC identifying individual from minute biological samples (tissue typing),  
 CC and in forensic identification of a biological sample. The present  
 CC sequence is a protein showing sequence similarity to a NOVX protein.  
 XX  
 XX Sequence 248 AA;

Query Match 73.7%; Score 1276; DB 7; Length 248;  
 Best Local Similarity 99.2%; Pred. No. 1.1e-115;  
 Matches 242; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 85 VFDEVDTPQSGVLSDRREITLARIHELPLSLQDLTGLEHMLNCSKMLPADITQNNIP 144  
 DB 5 VFDEVDTPQSGVLSDRREITLARIHELPLSLQDLTGLEHMLNCSKMLPADITQNNIP 64  
 QY 145 PTQESYDNPVPVTKSLVTNCKPVTDKIHRAYKDKNKYRFEIMGEEETAFKMIRTNVSH 204  
 DB 65 PTQESYDNPVPVTKSLVTNCKPVTDKIHRAYKDKNKYRFEIMGEEETAFKMIRTNVSH 124  
 QY 205 VVGQLDDTRKPKRFVCLNDINDHNHKAQAVLROFYFESMPFIPSQFELPRYRNF 264  
 DB 125 VVGQLDDTRKPKRFVCLNDINDHNHKAQAVLROFYFESMPFIPSQFELPRYRNF 184  
 QY 265 LHMHELOQWRAVRDKLKEWTHCVLATLIMFTIFSFABQIALKPKIPRRRIHKEASPN 324  
 DB 185 LHMHELOQWRAVRDKLKEWTHCVLATLIMFTIFSFABQIALKPKIPRRRIHKEASPN 244  
 QY 325 RIRV 328  
 DB 245 RIRV 248

RESULT 15  
 ADG39855  
 ID ADG39855 standard; protein; 663 AA.

XX AC ADG39855;  
 XX DT 26-FEB-2004 (first entry)  
 XX DE Protein similar to human NOV12 #1.  
 XX KW Human; NOVX; cardiomyopathy; atherosclerosis; hypertension;  
 XX KW congenital heart defect; pulmonary stenosis; scleroderma; obesity;  
 XX KW metabolic disturbance; obesity; transplantation; adrenoleukodystrophy;  
 XX KW congenital adrenal hyperplasia; prostate cancer; diabetes;  
 XX KW metabolic disorder; neoplasm; adenocarcinoma; fertility; haemophilia;  
 XX KW graft versus host disease; AIDS; bronchial asthma; Crohn's disease;  
 XX KW multiple sclerosis; infectious disease; anorexia;  
 XX KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
 XX KW immune disorder; haematopoietic disorder; dyslipidaemia;  
 XX KW wasting disorder.

XX OS Homo sapiens.  
 XX PN US2003203843-A1.  
 XX XX 30-OCT-2003.  
 XX PD 11-APR-2002; 2002US-00120801.  
 XX PF 20-APR-2001; 2001US-0285609P.  
 XX PR

PR 23-APR-2001; 2001US-0285748P.  
PR 24-APR-2001; 2001US-0285068P.  
PR 25-APR-2001; 2001US-0286292P.  
PR 03-MAY-2001; 2001US-0288334P.  
PR 16-MAY-2001; 2001US-0291241P.  
PR 14-SEP-2001; 2001US-0322284P.  
XX (PENA/) PENA C E A.  
PA (GUOX/) GUO X.  
PA (SHIM/) SHIMKETS R A.  
PA (PADI/) PADIGARU M.  
PA (KEKU/) KEKUDA R.  
PA (SPYT/) SPYTEK K A.  
PA (MEHR/) MEHRABAN F.  
PA (TOPP/) TOPPER J N.  
PA (WASS/) WASSERMAN U M.  
PA (NALLY/) NALLYANKAR S M.  
PA (EDIN/) EDINGER S R.  
PA (SMIT/) SMITHSON G.  
PA (GUNT/) GUNTHER E.  
PA (KOMU/) KOMUVES L.  
XX  
PI Pena CEA, Guo X, Shimkets RA, Padigaru M, Kekuda R, Spytek KA;  
PI Mehraban F, Topper JN, Malyankar UM, Wasserman SM, Edinger SR;  
PI Smithson G, Gunther E, Komuves L;  
XX WPI; 2003-900671/82.  
XX  
XX New NOVX polypeptides and nucleic acids, useful for diagnosing or  
PT treating e.g. cardiomyopathy, atherosclerosis, hypertension, scleroderma,  
PT obesity, prostate cancer, AIDS, bronchial asthma, Crohn's disease, or  
PT multiple sclerosis.  
XX  
PS Disclosure; SEQ ID NO 87; 215pp; English.  
XX  
XX The invention relates to a new isolated polypeptide comprising an amino  
CC acid sequence selected from 17 fully defined human NOVX sequences (even  
CC SEQ ID NOS between ADG39770 and ADG39802), a mature form of the NOVX  
CC amino acid or a variant of NOVX, where one or more amino acid residue in  
CC the variant differs in no more than 15% of the amino acid residues of  
CC NOVX. Also included are an isolated nucleic acid (NA) molecule  
CC (comprising a nucleic acid sequence encoding a NOVX polypeptide above  
CC (odd SEQ ID NOS between ADG39769 and ADG39801), a nucleic acid fragment  
CC encoding at least a portion of a NOVX polypeptide and a complement of  
CC NOVX NA), a vector comprising NOVX NA, a cell comprising the vector, an  
CC antibody that immunospecifically binds to NOVX, a method for determining  
CC the presence or amount of NOVX or NOVX NA in a sample, a method of  
CC identifying an agent that binds to NOVX, a method for identifying an  
CC agent that modulates the expression or activity of NOVX, a method for  
CC modulating the activity of NOVX, a method of treating or preventing a  
CC NOVX-associated disorder, a method for screening for a modulator of  
CC activity or of latency or predisposition to a NOVX-associated disorder, a  
CC method for determining the presence of or predisposition to a disease  
CC associated with altered levels of NOVX or NOVX NA in a first mammalian  
CC subject and a method of treating a pathological state in a mammal by  
CC administering NOVX or an antibody that binds to NOVX. The NOVX  
CC polypeptide, nucleic acid or antibody is useful in the manufacture of a  
CC medicament for treating a syndrome associated with a human disease or a  
CC NOVX-associated disorder. The NOVX polypeptides and nucleic acids  
CC encoding them are useful for diagnosing or treating pathologies, diseases  
CC or conditions associated with NOVX sequences, including cardiomyopathy,  
CC atherosclerosis, hypertension, congenital heart defects, pulmonary  
CC stenosis, scleroderma, obesity, metabolic disturbances associated with  
CC obesity, transplantation, adrenoleukodystrophy, congenital adrenal  
CC hyperplasia, prostate cancer, diabetes, metabolic disorders, neoplasm,  
CC adenocarcinoma, fertility, haemophilia, graft versus host disease, AIDS,  
CC bronchial asthma, Crohn's disease, multiple sclerosis, infectious  
CC disease, anorexia, neurodegenerative disorders (e.g. Alzheimer's disease,  
CC or Parkinson's disease), immune disorders, haematopoietic disorders,  
CC dyslipidaemias, and wasting disorders associated with chronic diseases.  
CC The polypeptides can be used as immunogens to produce antibodies and as  
CC vaccines. The sequences may further be used in chromosome mapping,  
CC identifying individual from minute biological samples (tissue typing),

CC and in forensic identification of a biological sample. The present  
CC sequence is a protein showing sequence similarity to a NOVX protein.  
XX  
SQ Sequence 663 AA;  
Query Match 65.9%; Score 1140; DB 7; Length 663;  
Best Local Similarity 98.6%; Pred. No. 8.3e-102; Indels 0; Gaps 0;  
Matches 218; Conservative 1; Mismatches 2;  
QY 1 DTFADSLRYVKNILNSKFGFTSRKVPAPHPMDIRVMQELQDMFPEEDKTSFHKVRHS 60  
DB 413 DTFADSLRYVKNILNSKFGFTSRKVPAPHPMDIRVMQELQDMFPEEDKTSFHKVRHS 472  
QY 61 EDMQFAFSFYFLMSAVOPLNISQVFEVDVTDOSGVLSDREIRTLATRIHELPLSLQDLT 120  
DB 473 EDMQFAFSFYFLMSAVOPLNISQVFEVDVTDOSGVLSDREIRTLATRIHELPLSLQDLT 532  
QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYYDPNLPVTKSLVTNCKPVTDKTHKAYKDK 180  
DB 533 GLEHMLINCSKMLPADITQLNNIPPTQESYYDPNLPVTKSLVTNCKPVTDKTHKAYKDK 592  
QY 181 NKYRFEIMGEEETAFKMRITNVSHVVGQLDDIRKNRKFVC 221  
DB 593 NKYRFEIMGEEETAFKMRITNVSHVVGQLDDIRKNRPRISLC 633

Search completed: November 21, 2004, 12:54:46

Job time : 39.6766 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2004, 12:47:18 ; Search time 9.88676 Seconds  
(without alignments)  
2200.144 Million cell updates/sec

Title: US-10-023-888-5  
Perfect score: 1731  
Sequence: 1 DTFADSLRYVYNKIILNSKFGF.....RKIFPRRIHKEASPNRIRV 328

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/pCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1731	100.0	328	4	US-09-635-872A-2
2	1731	100.0	328	4	US-09-636-077A-2
3	1731	100.0	328	4	US-09-636-060C-2
4	1731	100.0	328	4	US-09-986-552-2
5	1731	100.0	328	4	US-09-636-596C-2
6	1698	98.1	328	4	US-09-635-872A-8
7	1698	98.1	328	4	US-09-636-077A-8
8	1698	98.1	328	4	US-09-636-060C-8
9	1698	98.1	328	4	US-09-986-552-8
10	1698	98.1	328	4	US-09-636-596C-8
11	402	23.2	502	4	US-09-635-872A-13
12	402	23.2	502	4	US-09-636-077A-13
13	402	23.2	502	4	US-09-636-060C-13
14	402	23.2	502	4	US-09-986-552-13
15	402	23.2	502	4	US-09-636-596C-13
16	269	15.5	299	4	US-09-270-767-37559
17	269	15.5	299	4	US-09-270-767-52776
18	145	8.4	28	4	US-09-635-872A-27
19	145	8.4	28	4	US-09-636-077A-27
20	145	8.4	28	4	US-09-636-060C-27
21	145	8.4	28	4	US-09-986-552-27
22	145	8.4	28	4	US-09-636-596C-27
23	116	6.7	44	4	US-09-270-767-33278
24	116	6.7	44	4	US-09-270-767-48495
25	113	6.6	545	4	US-08-936-107A-10
26	101	5.8	2662	4	US-09-595-684B-31
27	101	5.8	2663	4	US-09-538-092-1252

28	100.5	5.8	259	3	US-08-469-318-137	Sequence 137, App
29	100.5	5.8	259	3	US-08-469-318-155	Sequence 155, App
30	100.5	5.8	259	3	US-08-468-609A-137	Sequence 137, App
31	100.5	5.8	259	3	US-08-468-609A-155	Sequence 155, App
32	100.5	5.8	259	3	US-08-446-872A-137	Sequence 137, App
33	100.5	5.8	259	3	US-08-446-872A-155	Sequence 155, App
34	100.5	5.8	259	4	US-08-762-227A-137	Sequence 137, App
35	100.5	5.8	259	4	US-08-762-227A-155	Sequence 155, App
36	100.5	5.8	259	5	PCT-US95-01185-137	Sequence 137, App
37	100.5	5.8	259	5	PCT-US95-01185-155	Sequence 155, App
38	96.5	5.6	588	4	US-09-519-232-2	Sequence 2, Appli
39	94	5.4	810	4	US-09-538-092-596	Sequence 596, App
40	93.5	5.4	652	3	US-08-559-896B-2	Sequence 2, Appli
41	93.5	5.4	652	4	US-09-351-794A-2	Sequence 2, Appli
42	93.5	5.4	811	4	US-09-248-796A-18641	Sequence 18641, A
43	93.5	5.4	1199	3	US-09-208-742-2	Sequence 2, Appli
44	93.5	5.4	1199	3	US-09-332-295-4	Sequence 4, Appli
45	93.5	5.4	1199	4	US-09-709-979-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1

US-09-635-872A-2  
; Sequence 2, Application US/09635872A  
; Patent No. 6534300  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASE;  
; FILE REFERENCE: 195613US0  
; CURRENT APPLICATION NUMBER: US/09/635,872A  
; CURRENT FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-635-872A-2

Query Match 100.0%; Score 1731; DB 4; Length 328;  
Best Local Similarity 100.0%; Pred. No. 1.5e-168;  
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	DTFADSLRYVYNKIILNSKFGFTSRKVPAPHPMHIDRIVMQLQDMPEPFDKTSFHKVRHS	60
DB	1	DTFADSLRYVYNKIILNSKFGFTSRKVPAPHPMHIDRIVMQLQDMPEPFDKTSFHKVRHS	60
QY	61	EDMQPAPSYFYLLMSAVQPLNISQVFDVDTQSGVLSDREIRTLATRIHELPLSLQDLT	120
DB	61	EDMQPAPSYFYLLMSAVQPLNISQVFDVDTQSGVLSDREIRTLATRIHELPLSLQDLT	120
QY	121	GLEHMLNCSKWLPAIDITOLANNIPPTQESYDPLPPVTKSLVTNCKEVTDKIHAYKDK	180
DB	121	GLEHMLNCSKWLPAIDITOLANNIPPTQESYDPLPPVTKSLVTNCKEVTDKIHAYKDK	180
QY	181	NKYRFEIINGEETIAFMIRTNVSHVVGQLDDIRKNPKRFVCLNDNIHNHDKDAQVTKAVL	240
DB	181	NKYRFEIINGEETIAFMIRTNVSHVVGQLDDIRKNPKRFVCLNDNIHNHDKDAQVTKAVL	240
QY	241	RDYFESMPPIPSQFELPREYRNRLHMEQLQEWAYRDKLFWTHCVLATLIMFTIPSF	300
DB	241	RDYFESMPPIPSQFELPREYRNRLHMEQLQEWAYRDKLFWTHCVLATLIMFTIPSF	300
QY	301	ABQLIALKRKIFPRRIHKEASPNRIRV	328
DB	301	ABQLIALKRKIFPRRIHKEASPNRIRV	328

RESULT 2

US-09-636-077A-2		US-09-636-077A-2	
; Sequence 2, Application US/09636077A		; Sequence 2, Application US/09636077A	
; Patent No. 6537785		; Patent No. 6537785	
; GENERAL INFORMATION:		; GENERAL INFORMATION:	
; APPLICANT: CANFIELD, WILLIAM		; APPLICANT: CANFIELD, WILLIAM	
; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE		; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE	
; FILE REFERENCE: 195612050		; FILE REFERENCE: 195612050	
; CURRENT APPLICATION NUMBER: US/09/636,077A		; CURRENT APPLICATION NUMBER: US/09/636,077A	
; CURRENT FILING DATE: 2000-08-10		; CURRENT FILING DATE: 2000-08-10	
; PRIOR APPLICATION NUMBER: 60/153,831		; PRIOR APPLICATION NUMBER: 60/153,831	
; PRIOR FILING DATE: 1999-09-14		; PRIOR FILING DATE: 1999-09-14	
; NUMBER OF SEQ ID NOS: 52		; NUMBER OF SEQ ID NOS: 52	
; SOFTWARE: Patentin version 3.0		; SOFTWARE: Patentin version 3.1	
; SEQ ID NO 2		; SEQ ID NO 2	
; LENGTH: 328		; LENGTH: 328	
; TYPE: PRT		; TYPE: PRT	
; ORGANISM: Homo sapiens		; ORGANISM: Homo sapiens	
US-09-636-077A-2		US-09-636-060C-2	
Query Match		Query Match	
Best Local Similarity 100.0%; Score 1731; DB 4; Length 328;		Best Local Similarity 100.0%; Score 1731; DB 4; Length 328;	
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	DTFADSLRYVKNILNSKFGFTSRKVPAPMPHMDIRIVMOELQDMFPEEDKTSFHKVRHS	60
DB	1	DTFADSLRYVKNILNSKFGFTSRKVPAPMPHMDIRIVMOELQDMFPEEDKTSFHKVRHS	60
QY	61	EDMQAFSYFYLLMSAVQPLNISQVDEVDVTDQSGVLSDRREITLARIHELPLSLQDLT	120
DB	61	EDMQAFSYFYLLMSAVQPLNISQVDEVDVTDQSGVLSDRREITLARIHELPLSLQDLT	120
QY	121	GLEHMLINCSKMLPADITQLNNIPPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYKDK	180
DB	121	GLEHMLINCSKMLPADITQLNNIPPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYKDK	180
QY	181	NKYRFEIMGEEIEAFKMRITNVSHVVGQLDDIRKNPKRFVCLNDNDHNDHDKDAQTVKAVL	240
DB	181	NKYRFEIMGEEIEAFKMRITNVSHVVGQLDDIRKNPKRFVCLNDNDHNDHDKDAQTVKAVL	240
QY	241	RDYESMFPPIPSQFELPREYRNFLHMHGELQEWAYRDKLFKFWTHCVLATLIMFTIFSFF	300
DB	241	RDYESMFPPIPSQFELPREYRNFLHMHGELQEWAYRDKLFKFWTHCVLATLIMFTIFSFF	300
QY	301	AEQIALKXKIFPRRRIHKEASPNRIIV	328
DB	301	AEQIALKXKIFPRRRIHKEASPNRIIV	328
RESULT 4		RESULT 3	
US-09-986-552-2		US-09-636-060C-2	
; Sequence 2, Application US/09986552		; Sequence 2, Application US/09636060C	
; Patent No. 6670165		; Patent No. 6642038	
; GENERAL INFORMATION:		; GENERAL INFORMATION:	
; APPLICANT: CANFIELD, WILLIAM		; APPLICANT: CANFIELD, WILLIAM M	
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES		; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY	
; FILE REFERENCE: 215089US77DIV		; FILE REFERENCE: 210119USOCNT	
; CURRENT APPLICATION NUMBER: US/09/986,552		; CURRENT APPLICATION NUMBER: US/09/636,060C	
; CURRENT FILING DATE: 2001-11-09		; CURRENT FILING DATE: 2000-08-10	
; PRIOR APPLICATION NUMBER: 09/635,872		; PRIOR APPLICATION NUMBER: 60/153,831	
; PRIOR FILING DATE: 2000-08-10		; PRIOR FILING DATE: 1999-09-14	
; PRIOR APPLICATION NUMBER: 60/153,831		; NUMBER OF SEQ ID NOS: 52	
; PRIOR FILING DATE: 1999-09-14		; SOFTWARE: Patentin version 3.1	
; NUMBER OF SEQ ID NOS: 52		; SEQ ID NO 2	
; SOFTWARE: Patentin version 3.1		; LENGTH: 328	
; SEQ ID NO 2		; TYPE: PRT	
; LENGTH: 328		; ORGANISM: Homo sapiens	
; TYPE: PRT		; ORGANISM: Homo sapiens	
; ORGANISM: Homo sapiens		US-09-636-060C-2	
US-09-986-552-2		Query Match	
Query Match		Best Local Similarity 100.0%; Score 1731; DB 4; Length 328;	
Best Local Similarity 100.0%; Pred. No. 1.5e-168;		Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	DTFADSLRYVKNILNSKFGFTSRKVPAPMPHMDIRIVMOELQDMFPEEDKTSFHKVRHS	60
DB	1	DTFADSLRYVKNILNSKFGFTSRKVPAPMPHMDIRIVMOELQDMFPEEDKTSFHKVRHS	60
QY	61	EDMQAFSYFYLLMSAVQPLNISQVDEVDVTDQSGVLSDRREITLARIHELPLSLQDLT	120
DB	61	EDMQAFSYFYLLMSAVQPLNISQVDEVDVTDQSGVLSDRREITLARIHELPLSLQDLT	120
QY	121	GLEHMLINCSKMLPADITQLNNIPPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYKDK	180
DB	121	GLEHMLINCSKMLPADITQLNNIPPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYKDK	180
QY	181	NKYRFEIMGEEIEAFKMRITNVSHVVGQLDDIRKNPKRFVCLNDNDHNDHDKDAQTVKAVL	240
DB	181	NKYRFEIMGEEIEAFKMRITNVSHVVGQLDDIRKNPKRFVCLNDNDHNDHDKDAQTVKAVL	240
QY	241	RDYESMFPPIPSQFELPREYRNFLHMHGELQEWAYRDKLFKFWTHCVLATLIMFTIFSFF	300
DB	241	RDYESMFPPIPSQFELPREYRNFLHMHGELQEWAYRDKLFKFWTHCVLATLIMFTIFSFF	300
QY	301	AEQIALKXKIFPRRRIHKEASPNRIIV	328
DB	301	AEQIALKXKIFPRRRIHKEASPNRIIV	328
Query Match		Query Match	
Best Local Similarity 100.0%; Score 1731; DB 4; Length 328;		Best Local Similarity 100.0%; Score 1731; DB 4; Length 328;	
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	DTFADSLRYVKNILNSKFGFTSRKVPAPMPHMDIRIVMOELQDMFPEEDKTSFHKVRHS	60
DB	1	DTFADSLRYVKNILNSKFGFTSRKVPAPMPHMDIRIVMOELQDMFPEEDKTSFHKVRHS	60
QY	61	EDMQAFSYFYLLMSAVQPLNISQVDEVDVTDQSGVLSDRREITLARIHELPLSLQDLT	120
DB	61	EDMQAFSYFYLLMSAVQPLNISQVDEVDVTDQSGVLSDRREITLARIHELPLSLQDLT	120
QY	121	GLEHMLINCSKMLPADITQLNNIPPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYKDK	180
DB	121	GLEHMLINCSKMLPADITQLNNIPPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYKDK	180
QY	181	NKYRFEIMGEEIEAFKMRITNVSHVVGQLDDIRKNPKRFVCLNDNDHNDHDKDAQTVKAVL	240
DB	181	NKYRFEIMGEEIEAFKMRITNVSHVVGQLDDIRKNPKRFVCLNDNDHNDHDKDAQTVKAVL	240
QY	241	RDYESMFPPIPSQFELPREYRNFLHMHGELQEWAYRDKLFKFWTHCVLATLIMFTIFSFF	300
DB	241	RDYESMFPPIPSQFELPREYRNFLHMHGELQEWAYRDKLFKFWTHCVLATLIMFTIFSFF	300
QY	301	AEQIALKXKIFPRRRIHKEASPNRIIV	328
DB	301	AEQIALKXKIFPRRRIHKEASPNRIIV	328
Query Match		Query Match	
Best Local Similarity 100.0%; Score 1731; DB 4; Length 328;		Best Local Similarity 100.0%; Score 1731; DB 4; Length 328;	
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	DTFADSLRYVKNILNSKFGFTSRKVPAPMPHMDIRIVMOELQDMFPEEDKTSFHKVRHS	60
DB	1	DTFADSLRYVKNILNSKFGFTSRKVPAPMPHMDIRIVMOELQDMFPEEDKTSFHKVRHS	60
QY	61	EDMQAFSYFYLLMSAVQPLNISQVDEVDVTDQSGVLSDRREITLARIHELPLSLQDLT	120
DB	61	EDMQAFSYFYLLMSAVQPLNISQVDEVDVTDQSGVLSDRREITLARIHELPLSLQDLT	120
QY	121	GLEHMLINCSKMLPADITQLNNIPPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYKDK	180
DB	121	GLEHMLINCSKMLPADITQLNNIPPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYKDK	180
QY	181	NKYRFEIMGEEIEAFKMRITNVSHVVGQLDDIRKNPKRFVCLNDNDHNDHDKDAQTVKAVL	240
DB	181	NKYRFEIMGEEIEAFKMRITNVSHVVGQLDDIRKNPKRFVCLNDNDHNDHDKDAQTVKAVL	240
QY	241	RDYESMFPPIPSQFELPREYRNFLHMHGELQEWAYRDKLFKFWTHCVLATLIMFTIFSFF	300
DB	241	RDYESMFPPIPSQFELPREYRNFLHMHGELQEWAYRDKLFKFWTHCVLATLIMFTIFSFF	300
QY	301	AEQIALKXKIFPRRRIHKEASPNRIIV	328
DB	301	AEQIALKXKIFPRRRIHKEASPNRIIV	328

```
RESULT 5
US-09-636-596C-2
; Sequence 2, Application US/09636596C
; Patent No. 6770468
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: PHOSPHODIESTER ALPHA-GlcNacase OF THE LYSSOSOMAL TARGETING PATHWAY
; FILE REFERENCE: 10929-0001-77
; CURRENT APPLICATION NUMBER: US/09/636.596C
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-636-596C-2

Query Match      100.0%; Score 1731; DB 4; Length 328;
Best Local Similarity 100.0%; Pred. No. 1.5e-168;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTFADSLRYVVKILNSKFGFTSRKVPAAHMPHMDIRIVMQLQDMFPEEDKTSFHKVRS 60
Db 1 DTFADSLRYVVKILNSKFGFTSRKVPAAHMPHMDIRIVMQLQDMFPEEDKTSFHKVRS 60
Qy 61 EDMQAFSYFYILMSAVOPLNISQVFEVDVTDQSGVLSDEIRTLATRIHELPLSLQDLT 120
Db 61 EDMQAFSYFYILMSAVOPLNISQVFEVDVTDQSGVLSDEIRTLATRIHELPLSLQDLT 120
Qy 121 GLEHMLINCSKMLPADITQLANNIPTQESYYDPNLPVTKSLVTNCKPVTDKIHKAYKDK 180
Db 121 GLEHMLINCSKMLPADITQLANNIPTQESYYDPNLPVTKSLVTNCKPVTDKIHKAYKDK 180
Qy 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPRKFVCLNDNDHNHKAQTVKAVL 240
Db 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPRKFVCLNDNDHNHKAQTVKAVL 240
Qy 241 RDPYESMFPPIPSQFELPREYRNRFLEHMQEWRAYRDKLKFTHCVLATLIMFTIFSFF 300
Db 241 RDPYESMFPPIPSQFELPREYRNRFLEHMQEWRAYRDKLKFTHCVLATLIMFTIFSFF 300
Qy 301 AEQIALKPKIFPRRIHKEASPNRIRV 328
Db 301 AEQIALKPKIFPRRIHKEASPNRIRV 328

RESULT 6
US-09-635-872A-8
; Sequence 8, Application US/09635872A
; Patent No. 6534300
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSSOSOMAL HYDROLASES
; FILE REFERENCE: 195613USO
; CURRENT APPLICATION NUMBER: US/09/635.872A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 8
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-635-872A-8

Query Match      98.1%; Score 1698; DB 4; Length 328;
Best Local Similarity 97.6%; Pred. No. 3.4e-165;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DTFADSLRYVVKILNSKFGFTSRKVPAAHMPHMDIRIVMQLQDMFPEEDKTSFHKVRS 60
Db 1 DTFADSLRYVVKILNSKFGFTSRKVPAAHMPHMDIRIVMQLQDMFPEEDKTSFHKVRS 60
Qy 61 EDMQAFSYFYILMSAVOPLNISQVFEVDVTDQSGVLSDEIRTLATRIHELPLSLQDLT 120
Db 61 EDMQAFSYFYILMSAVOPLNISQVFEVDVTDQSGVLSDEIRTLATRIHELPLSLQDLT 120
Qy 121 GLEHMLINCSKMLPADITQLANNIPTQESYYDPNLPVTKSLVTNCKPVTDKIHKAYKDK 180
Db 121 GLEHMLINCSKMLPADITQLANNIPTQESYYDPNLPVTKSLVTNCKPVTDKIHKAYKDK 180
Qy 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPRKFVCLNDNDHNHKAQTVKAVL 240
Db 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPRKFVCLNDNDHNHKAQTVKAVL 240
Qy 241 RDPYESMFPPIPSQFELPREYRNRFLEHMQEWRAYRDKLKFTHCVLATLIMFTIFSFF 300
Db 241 RDPYESMFPPIPSQFELPREYRNRFLEHMQEWRAYRDKLKFTHCVLATLIMFTIFSFF 300
Qy 301 AEQIALKPKIFPRRIHKEASPNRIRV 328
Db 301 AEQIALKPKIFPRRIHKEASPNRIRV 328

RESULT 7
US-09-636-077A-8
; Sequence 8, Application US/09636077A
; Patent No. 6537785
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS OF TREATING LYSSOSOMAL STORAGE DISEASE
; FILE REFERENCE: 195612USO
; CURRENT APPLICATION NUMBER: US/09/636.077A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 8
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-636-077A-8

Query Match      98.1%; Score 1698; DB 4; Length 328;
Best Local Similarity 97.6%; Pred. No. 3.4e-165;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DTFADSLRYVVKILNSKFGFTSRKVPAAHMPHMDIRIVMQLQDMFPEEDKTSFHKVRS 60
Db 1 DTFADSLRYVVKILNSKFGFTSRKVPAAHMPHMDIRIVMQLQDMFPEEDKTSFHKVRS 60
Qy 61 EDMQAFSYFYILMSAVOPLNISQVFEVDVTDQSGVLSDEIRTLATRIHELPLSLQDLT 120
Db 61 EDMQAFSYFYILMSAVOPLNISQVFEVDVTDQSGVLSDEIRTLATRIHELPLSLQDLT 120
Qy 121 GLEHMLINCSKMLPADITQLANNIPTQESYYDPNLPVTKSLVTNCKPVTDKIHKAYKDK 180
Db 121 GLEHMLINCSKMLPADITQLANNIPTQESYYDPNLPVTKSLVTNCKPVTDKIHKAYKDK 180
Qy 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPRKFVCLNDNDHNHKAQTVKAVL 240
Db 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQDDIRKNPRKFVCLNDNDHNHKAQTVKAVL 240
Qy 241 RDPYESMFPPIPSQFELPREYRNRFLEHMQEWRAYRDKLKFTHCVLATLIMFTIFSFF 300
Db 241 RDPYESMFPPIPSQFELPREYRNRFLEHMQEWRAYRDKLKFTHCVLATLIMFTIFSFF 300
Qy 301 AEQIALKPKIFPRRIHKEASPNRIRV 328
Db 301 AEQIALKPKIFPRRIHKEASPNRIRV 328
```

Db 301 AEQIIALKRKIFPERRIHKEASPDRIYV 328

## RESULT 8

US-09-636-060C-8  
; Sequence 8, Application US/09636060C  
; Patent No. 6642038

## GENERAL INFORMATION:

; APPLICANT: CANFIELD, WILLIAM M  
; TITLE OF INVENTION: GLCNAC PHOSPHOTRANSFERASE OF THE LYSOSOMAL TARGETING PATHWAY  
; FILE REFERENCE: 2101-9USOCONT  
; CURRENT APPLICATION NUMBER: US/09/636,060C  
; CURRENT FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 8  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-636-060C-8

Query Match 98.1%; Score 1698; DB 4; Length 328;

Best Local Similarity 97.6%; Pred. No. 3.4e-165;  
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVVKILNSKFGFTSRKVPAPMPHMDRIVMQELQDMFPEEFDKTSFHKYRHS 60

Db 1 DTFADSLRYVVKILNSKFGFTSRKVPAPMPHMDRIVMQELQDMFPEEFDKTSFHKYRHS 60

QY 61 EDMQFAFSFYLLMSAVOPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHDLPLSLQDLT 120

Db 61 EDMQFAFSFYLLMSAVOPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHDLPLSLQDLT 120

QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYVDNLPVTKSLVTKNCKPVTDKIHKAYKDK 180

Db 121 GLEHMLINCSKMLPANITQLNNIPPTQEAAYDNLPPVTKSLVTKNCKPVTDKIHKAYKDK 180

QY 181 NKYRFEIMGEEBEEIAFKMIRTNVSHVVGQDDIRKNPKRFVCLNDNDHNDHKAQTVKAVL 240

Db 181 NKYRFEIMGEEBEEIAFKMIRTNVSHVVGQDDIRKNPKRFVCLNDNDHNDHKAQTVKAVL 240

QY 241 RDFVESMPPIPSQPELPREYRNFLHMHLEQWRAVDKLFKFWTHCVLATLIMFTIFSFF 300

Db 241 RDFVESMPPIPSQPELPREYRNFLHMHLEQWRAVDKLFKFWTHCVLATLIMFTIFSFF 300

QY 301 AEQIIALKRKIFPERRIHKEASPDRIYV 328

Db 301 AEQIIALKRKIFPERRIHKEASPDRIYV 328

## RESULT 9

US-09-986-552-8  
; Sequence 8, Application US/09986552  
; Patent No. 6670165

## GENERAL INFORMATION:

; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 215089US77DIV  
; CURRENT APPLICATION NUMBER: US/09/986,552  
; CURRENT FILING DATE: 2001-11-09  
; PRIOR APPLICATION NUMBER: 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 8  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-986-552-8

Query Match 98.1%; Score 1698; DB 4; Length 328;

Best Local Similarity 97.6%; Pred. No. 3.4e-165;  
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVVKILNSKFGFTSRKVPAPMPHMDRIVMQELQDMFPEEFDKTSFHKYRHS 60

Db 1 DTFADSLRYVVKILNSKFGFTSRKVPAPMPHMDRIVMQELQDMFPEEFDKTSFHKYRHS 60

QY 61 EDMQFAFSFYLLMSAVOPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHDLPLSLQDLT 120

Db 61 EDMQFAFSFYLLMSAVOPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHDLPLSLQDLT 120

QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYVDNLPVTKSLVTKNCKPVTDKIHKAYKDK 180

Db 121 GLEHMLINCSKMLPANITQLNNIPPTQEAAYDNLPPVTKSLVTKNCKPVTDKIHKAYKDK 180

QY 181 NKYRFEIMGEEBEEIAFKMIRTNVSHVVGQDDIRKNPKRFVCLNDNDHNDHKAQTVKAVL 240

Db 181 NKYRFEIMGEEBEEIAFKMIRTNVSHVVGQDDIRKNPKRFVCLNDNDHNDHKAQTVKAVL 240

QY 241 RDFVESMPPIPSQPELPREYRNFLHMHLEQWRAVDKLFKFWTHCVLATLIMFTIFSFF 300

Db 241 RDFVESMPPIPSQPELPREYRNFLHMHLEQWRAVDKLFKFWTHCVLATLIMFTIFSFF 300

QY 301 AEQIIALKRKIFPERRIHKEASPDRIYV 328

Db 301 AEQIIALKRKIFPERRIHKEASPDRIYV 328

## RESULT 10

US-09-636-596C-8

; Sequence 8, Application US/09636596C  
; Patent No. 6770468

## GENERAL INFORMATION:

; APPLICANT: CANFIELD, WILLIAM  
; TITLE OF INVENTION: PHOSPHODIESTER ALPHA-GLCNACase OF THE LYSOSOMAL TARGETING PATHWAY  
; FILE REFERENCE: 10929-0001-77  
; CURRENT APPLICATION NUMBER: US/09/636,596C  
; CURRENT FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 8  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-636-596C-8

Query Match 98.1%; Score 1698; DB 4; Length 328;

Best Local Similarity 97.6%; Pred. No. 3.4e-165;  
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVVKILNSKFGFTSRKVPAPMPHMDRIVMQELQDMFPEEFDKTSFHKYRHS 60

Db 1 DTFADSLRYVVKILNSKFGFTSRKVPAPMPHMDRIVMQELQDMFPEEFDKTSFHKYRHS 60

QY 61 EDMQFAFSFYLLMSAVOPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHDLPLSLQDLT 120

Db 61 EDMQFAFSFYLLMSAVOPLNISQVFEVDVTDQSGVLSDBREIRTLATRIHDLPLSLQDLT 120

QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYVDNLPVTKSLVTKNCKPVTDKIHKAYKDK 180

Db 121 GLEHMLINCSKMLPANITQLNNIPPTQEAAYDNLPPVTKSLVTKNCKPVTDKIHKAYKDK 180

QY 181 NKYRFEIMGEEBEEIAFKMIRTNVSHVVGQDDIRKNPKRFVCLNDNDHNDHKAQTVKAVL 240

Db 181 NKYRFEIMGEEBEEIAFKMIRTNVSHVVGQDDIRKNPKRFVCLNDNDHNDHKAQTVKAVL 240

QY 241 RDFVESMPPIPSQPELPREYRNFLHMHLEQWRAVDKLFKFWTHCVLATLIMFTIFSFF 300

Db 241 RDFVESMPPIPSQPELPREYRNFLHMHLEQWRAVDKLFKFWTHCVLATLIMFTIFSFF 300







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2004, 12:56:09 ; Search time 35.4053 Seconds  
(without alignments)  
3280.691 Million cell updates/sec

Title: US-10-023-888-5  
Perfect score: 1731  
Sequence: 1 DTFADSLRYVVKILNSKPGF.....RKIPFRRRIHKEASPNRIRV 328

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1570615 seqs, 354127592 residues

Total number of hits satisfying chosen parameters: 1570615

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Published Applications AA:\*
- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pap.\*
  - 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pap.\*
  - 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pap.\*
  - 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pap.\*
  - 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pap.\*
  - 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pap.\*
  - 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pap.\*
  - 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pap.\*
  - 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pap.\*
  - 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pap.\*
  - 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pap.\*
  - 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pap.\*
  - 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pap.\*
  - 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pap.\*
  - 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pap.\*
  - 16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pap.\*
  - 17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pap.\*
  - 18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pap.\*
  - 19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pap.\*
  - 20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1731	100.0	328	9	US-09-895-072-2
2	1731	100.0	328	9	US-09-986-552-2
3	1731	100.0	328	14	US-10-023-888-5
4	1731	100.0	328	14	US-10-023-889-5
5	1731	100.0	328	14	US-10-023-890-5
6	1731	100.0	328	14	US-10-024-197-5
7	1731	100.0	328	14	US-10-023-894-5
8	1731	100.0	328	14	US-10-023-895-2
9	1698	98.1	328	9	US-09-895-072-8
10	1698	98.1	328	9	US-09-986-552-8
11	1698	98.1	328	14	US-10-023-888-10
12	1698	98.1	328	14	US-10-023-889-10
13	1698	98.1	328	14	US-10-023-890-10

14	1698	98.1	328	14	US-10-024-197-10	Sequence 10, Appl
15	1698	98.1	328	14	US-10-023-894-10	Sequence 10, Appl
16	1698	98.1	328	14	US-10-023-886-8	Sequence 8, Appl
17	1487	85.9	1199	14	US-10-023-888-2	Sequence 2, Appl
18	1487	85.9	1199	14	US-10-023-889-2	Sequence 2, Appl
19	1487	85.9	1199	14	US-10-023-890-2	Sequence 2, Appl
20	1487	85.9	1199	14	US-10-024-197-2	Sequence 2, Appl
21	1487	85.9	1199	14	US-10-023-894-2	Sequence 2, Appl
22	1290	74.5	248	14	US-10-120-801-90	Sequence 90, Appl
23	1140	65.9	663	14	US-10-120-801-87	Sequence 87, Appl
24	1140	65.9	1459	14	US-10-120-801-26	Sequence 26, Appl
25	581	33.6	384	14	US-10-120-801-89	Sequence 89, Appl
26	506	29.2	652	14	US-10-120-801-91	Sequence 91, Appl
27	402	23.2	502	9	US-09-895-072-13	Sequence 13, Appl
28	402	23.2	502	9	US-09-986-552-13	Sequence 13, Appl
29	402	23.2	502	14	US-10-023-888-16	Sequence 16, Appl
30	402	23.2	502	14	US-10-023-889-16	Sequence 16, Appl
31	402	23.2	502	14	US-10-023-890-16	Sequence 16, Appl
32	402	23.2	502	14	US-10-024-197-16	Sequence 16, Appl
33	402	23.2	502	14	US-10-023-894-16	Sequence 16, Appl
34	402	23.2	502	14	US-10-023-886-13	Sequence 13, Appl
35	300	17.3	55	14	US-10-029-386-30775	Sequence 30775, A
36	228	13.2	112	13	US-10-079-623-349	Sequence 349, Appl
37	145	8.4	28	9	US-09-895-072-27	Sequence 27, Appl
38	145	8.4	28	9	US-09-986-552-27	Sequence 27, Appl
39	145	8.4	28	14	US-10-023-886-27	Sequence 27, Appl
40	124	7.2	1312	14	US-10-369-493-1950	Sequence 1950, Ap
41	112	6.5	46	14	US-10-023-888-26	Sequence 26, Appl
42	110.5	6.4	1015	16	US-10-437-963-160369	Sequence 160369,
43	104.5	6.0	937	16	US-10-437-963-156445	Sequence 156445,
44	104.5	6.0	1999	14	US-10-028-248A-107	Sequence 107, App
45	104.5	6.0	1999	15	US-10-107-782-107	Sequence 107, App

ALIGNMENTS

RESULT 1  
US-09-895-072-2  
; Sequence 2, Application US/09895072  
; Patent No. US2002025550A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM M  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 210119050CONT  
; CURRENT APPLICATION NUMBER: US/09/895,072  
; CURRENT FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; PRIOR APPLICATION NUMBER: US 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 2  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-895-072-2

Query Match 100.0%; Score 1731; DB 9; Length 328;  
Best Local Similarity 100.0%; Pred. No. 4.3e-149;  
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 DTFADSLRYVVKILNSKPGFTSRKYPAAHMPHMDIRVMQELQDMPEEFDDKTSFHKVRRHS 60  
DB 1 DTFADSLRYVVKILNSKPGFTSRKYPAAHMPHMDIRVMQELQDMPEEFDDKTSFHKVRRHS 60  
  
QY 61 EDMQAFSFFYYLMSAVQPLNISQVDFDEVDTQSGVLSDRERTLATRIHELPLSLQDIT 120  
DB 61 EDMQAFSFFYYLMSAVQPLNISQVDFDEVDTQSGVLSDRERTLATRIHELPLSLQDIT 120  
  
QY 121 GLEHMLNCSKMLPADITQLNNIPQSYDNPVPTKSLVTNCKVPTDKIHAYKDK 180

Db 121 GLEHMLNCSKMLPADITQLNNIPPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYDK 180  
Qy 181 NKYRFEIMGEIEAFKIRTNVSHVQGLDDIRKNPKFVCLNDNDHNDKDAQTAVKAVL 240  
Db 181 NKYRFEIMGEIEAFKIRTNVSHVQGLDDIRKNPKFVCLNDNDHNDKDAQTAVKAVL 240  
Qy 241 RDPYESMPPIPSQFELPREYRNFLHMHQEWRAVDKLFKFWTHCVLATLIMFTISFF 300  
Db 241 RDPYESMPPIPSQFELPREYRNFLHMHQEWRAVDKLFKFWTHCVLATLIMFTISFF 300  
Qy 301 AEQIALKXKIFPRRRHKEASPNRIRV 328  
Db 301 AEQIALKXKIFPRRRHKEASPNRIRV 328  
RESULT 2  
US-09-986-552-2  
; Sequence 2, Application US/05986552  
; Patent No. US20020150381A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 215089US77DIV  
; CURRENT APPLICATION NUMBER: US/09/986,552  
; CURRENT FILING DATE: 2001-11-09  
; PRIOR APPLICATION NUMBER: 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-986-552-2  
Query Match 100.0%; Score 1731; DB 9; Length 328;  
Best Local Similarity 100.0%; Pred. No. 4.3e-149;  
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 DTFADSLRYNKNILNSKFGFTSRKVPAPMHPMDIRIVMQELODMFPPEFDKTSFKVYRHS 60  
Db 1 DTFADSLRYNKNILNSKFGFTSRKVPAPMHPMDIRIVMQELODMFPPEFDKTSFKVYRHS 60  
Qy 61 EDMQAFSFFYFLMSAVQPLNISQVFEVDVTDQSGVLSDEIRTLATRIHELPLSLQDLT 120  
Db 61 EDMQAFSFFYFLMSAVQPLNISQVFEVDVTDQSGVLSDEIRTLATRIHELPLSLQDLT 120  
Qy 121 GLEHMLNCSKMLPADITQLNNIPPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYDK 180  
Db 121 GLEHMLNCSKMLPADITQLNNIPPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYDK 180  
Qy 181 NKYRFEIMGEIEAFKIRTNVSHVQGLDDIRKNPKFVCLNDNDHNDKDAQTAVKAVL 240  
Db 181 NKYRFEIMGEIEAFKIRTNVSHVQGLDDIRKNPKFVCLNDNDHNDKDAQTAVKAVL 240  
Qy 241 RDPYESMPPIPSQFELPREYRNFLHMHQEWRAVDKLFKFWTHCVLATLIMFTISFF 300  
Db 241 RDPYESMPPIPSQFELPREYRNFLHMHQEWRAVDKLFKFWTHCVLATLIMFTISFF 300  
Qy 301 AEQIALKXKIFPRRRHKEASPNRIRV 328  
Db 301 AEQIALKXKIFPRRRHKEASPNRIRV 328  
RESULT 3  
US-10-023-888-5  
; Sequence 5, Application US/10023888  
; Publication No. US20030119088A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: SOLUBLE GLCNAC PHOSPHOTRANSFERASE

; FILE REFERENCE: 203515US77  
; CURRENT APPLICATION NUMBER: US/10/023,888  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-023-888-5  
Query Match 100.0%; Score 1731; DB 14; Length 328;  
Best Local Similarity 100.0%; Pred. No. 4.3e-149;  
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 DTFADSLRYNKNILNSKFGFTSRKVPAPMHPMDIRIVMQELODMFPPEFDKTSFKVYRHS 60  
Db 1 DTFADSLRYNKNILNSKFGFTSRKVPAPMHPMDIRIVMQELODMFPPEFDKTSFKVYRHS 60  
Qy 61 EDMQAFSFFYFLMSAVQPLNISQVFEVDVTDQSGVLSDEIRTLATRIHELPLSLQDLT 120  
Db 61 EDMQAFSFFYFLMSAVQPLNISQVFEVDVTDQSGVLSDEIRTLATRIHELPLSLQDLT 120  
Qy 121 GLEHMLNCSKMLPADITQLNNIPPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYDK 180  
Db 121 GLEHMLNCSKMLPADITQLNNIPPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYDK 180  
Qy 181 NKYRFEIMGEIEAFKIRTNVSHVQGLDDIRKNPKFVCLNDNDHNDKDAQTAVKAVL 240  
Db 181 NKYRFEIMGEIEAFKIRTNVSHVQGLDDIRKNPKFVCLNDNDHNDKDAQTAVKAVL 240  
Qy 241 RDPYESMPPIPSQFELPREYRNFLHMHQEWRAVDKLFKFWTHCVLATLIMFTISFF 300  
Db 241 RDPYESMPPIPSQFELPREYRNFLHMHQEWRAVDKLFKFWTHCVLATLIMFTISFF 300  
Qy 301 AEQIALKXKIFPRRRHKEASPNRIRV 328  
Db 301 AEQIALKXKIFPRRRHKEASPNRIRV 328  
RESULT 4  
US-10-023-889-5  
; Sequence 5, Application US/10023889  
; Publication No. US20030124652A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS OF PRODUCING HIGH MANNOSE GLYCOPROTEINS IN COMPLEX CARBO  
; TITLE OF INVENTION: DEFICIENT CELLS  
; FILE REFERENCE: 203512US77  
; CURRENT APPLICATION NUMBER: US/10/023,889  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-023-889-5  
Query Match 100.0%; Score 1731; DB 14; Length 328;  
Best Local Similarity 100.0%; Pred. No. 4.3e-149;  
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 DTFADSLRYNKNILNSKFGFTSRKVPAPMHPMDIRIVMQELODMFPPEFDKTSFKVYRHS 60  
Db 1 DTFADSLRYNKNILNSKFGFTSRKVPAPMHPMDIRIVMQELODMFPPEFDKTSFKVYRHS 60  
Qy 61 EDMQAFSFFYFLMSAVQPLNISQVFEVDVTDQSGVLSDEIRTLATRIHELPLSLQDLT 120  
Db 61 EDMQAFSFFYFLMSAVQPLNISQVFEVDVTDQSGVLSDEIRTLATRIHELPLSLQDLT 120  
Qy 121 GLEHMLNCSKMLPADITQLNNIPPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYDK 180  
Db 121 GLEHMLNCSKMLPADITQLNNIPPTQESYYDNLPPVTKSLVTNCKPVTDKIHKAYDK 180

Db 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDPLPPVTKSLVTNCKPVTDKIHKAYKDK 180  
Qy 181 NKYRFEIMGEETIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNDHNHDKDAQTVKAVL 240  
Db 181 NKYRFEIMGEETIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNDHNHDKDAQTVKAVL 240  
Qy 241 RDFSMPPIPSQFELPREYRNRFHMHLEQWRAIRDKLKFVTHCVLATLIMFTISFF 300  
Db 241 RDFSMPPIPSQFELPREYRNRFHMHLEQWRAIRDKLKFVTHCVLATLIMFTISFF 300  
Qy 301 AEQIALKXKIFPRRIHKEASPNRIRV 328  
Db 301 AEQIALKXKIFPRRIHKEASPNRIRV 328

RESULT 5  
US-10-023-890-5  
; Sequence 5, Application US/10023890  
; Publication No. US20030124653A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHOD OF PRODUCING GLYCOPROTEINS HAVING REDUCED COMPLEX CARBOHYDRATE  
; FILE REFERENCE: 203510US77  
; CURRENT APPLICATION NUMBER: US/10/023,890  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-023-890-5

Query Match 100.0%; Score 1731; DB 14; Length 328;  
Best Local Similarity 100.0%; Pred. No. 4.3e-149;  
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPHPMHMIDRIVMOELQDMFPEEDKTSFHKVRHS 60  
Db 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPHPMHMIDRIVMOELQDMFPEEDKTSFHKVRHS 60  
Qy 61 EDMQFAFSFYFYLMSAVQPLNISQVDEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120  
Db 61 EDMQFAFSFYFYLMSAVQPLNISQVDEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120  
Qy 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDPLPPVTKSLVTNCKPVTDKIHKAYKDK 180  
Db 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDPLPPVTKSLVTNCKPVTDKIHKAYKDK 180  
Qy 181 NKYRFEIMGEETIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNDHNHDKDAQTVKAVL 240  
Db 181 NKYRFEIMGEETIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNDHNHDKDAQTVKAVL 240  
Qy 241 RDFSMPPIPSQFELPREYRNRFHMHLEQWRAIRDKLKFVTHCVLATLIMFTISFF 300  
Db 241 RDFSMPPIPSQFELPREYRNRFHMHLEQWRAIRDKLKFVTHCVLATLIMFTISFF 300  
Qy 301 AEQIALKXKIFPRRIHKEASPNRIRV 328  
Db 301 AEQIALKXKIFPRRIHKEASPNRIRV 328

RESULT 6  
US-10-024-197-5  
; Sequence 5, Application US/10024197  
; Publication No. US20030133924A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: HIGHLY PHOSPHORYLATED ACID BETA-GLUCOCEREOSIDASE AND METHODS OF  
; FILE REFERENCE: 209794USO  
; CURRENT APPLICATION NUMBER: US/10/024,197

; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-024-197-5

Query Match 100.0%; Score 1731; DB 14; Length 328;  
Best Local Similarity 100.0%; Pred. No. 4.3e-149;  
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPHPMHMIDRIVMOELQDMFPEEDKTSFHKVRHS 60  
Db 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPHPMHMIDRIVMOELQDMFPEEDKTSFHKVRHS 60  
Qy 61 EDMQFAFSFYFYLMSAVQPLNISQVDEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120  
Db 61 EDMQFAFSFYFYLMSAVQPLNISQVDEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120  
Qy 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDPLPPVTKSLVTNCKPVTDKIHKAYKDK 180  
Db 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDPLPPVTKSLVTNCKPVTDKIHKAYKDK 180  
Qy 181 NKYRFEIMGEETIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNDHNHDKDAQTVKAVL 240  
Db 181 NKYRFEIMGEETIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNDHNHDKDAQTVKAVL 240  
Qy 241 RDFSMPPIPSQFELPREYRNRFHMHLEQWRAIRDKLKFVTHCVLATLIMFTISFF 300  
Db 241 RDFSMPPIPSQFELPREYRNRFHMHLEQWRAIRDKLKFVTHCVLATLIMFTISFF 300  
Qy 301 AEQIALKXKIFPRRIHKEASPNRIRV 328  
Db 301 AEQIALKXKIFPRRIHKEASPNRIRV 328

## RESULT 7

US-10-023-894-5  
; Sequence 5, Application US/10023894  
; Publication No. US20030143669A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; APPLICANT: KORNFIELD, Stuart  
; TITLE OF INVENTION: EXPRESSION OF LYXOSOMAL HYDROLASE IN CELLS EXPRESSING PRO-N-  
; FILE REFERENCE: 217139US77  
; CURRENT APPLICATION NUMBER: US/10/023,894  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-023-894-5

Query Match 100.0%; Score 1731; DB 14; Length 328;  
Best Local Similarity 100.0%; Pred. No. 4.3e-149;  
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPHPMHMIDRIVMOELQDMFPEEDKTSFHKVRHS 60  
Db 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPHPMHMIDRIVMOELQDMFPEEDKTSFHKVRHS 60  
Qy 61 EDMQFAFSFYFYLMSAVQPLNISQVDEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120  
Db 61 EDMQFAFSFYFYLMSAVQPLNISQVDEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120  
Qy 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDPLPPVTKSLVTNCKPVTDKIHKAYKDK 180  
Db 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDPLPPVTKSLVTNCKPVTDKIHKAYKDK 180

QY 181 NKYFEIMGEETAFKMTNVSHVVGOLDIRKNPKFVCLNDNIDHNHKAQTVKAVL 240  
 DB 181 NKYFEIMGEETAFKMTNVSHVVGOLDIRKNPKFVCLNDNIDHNHKAQTVKAVL 240  
 QY 241 RDFSMPFIPSPQELPREYRNFLHMHQLQEWAYRDKLFWTHCVLATLIMFTIFSFF 300  
 DB 241 RDFSMPFIPSPQELPREYRNFLHMHQLQEWAYRDKLFWTHCVLATLIMFTIFSFF 300  
 QY 301 AEQIALKRRKIFPRRRHKEASPNRIRV 328  
 DB 301 AEQIALKRRKIFPRRRHKEASPNRIRV 328

RESULT 8  
 US-10-306-686-2  
 ; Sequence 2, Application US/10306686  
 ; Publication No. US20030148460A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CANFIELD, WILLIAM  
 ; TITLE OF INVENTION: PHOSPHODIESTER ALPHA-GLUCANASE OF THE LYSOSOMAL TARGETING PATHWAY  
 ; FILE REFERENCE: 230397US77DIV  
 ; CURRENT APPLICATION NUMBER: US/10/306,686  
 ; CURRENT FILING DATE: 2002-11-29  
 ; PRIOR APPLICATION NUMBER: 09/636,596  
 ; PRIOR FILING DATE: 2000-08-10  
 ; PRIOR APPLICATION NUMBER: 60/153,831  
 ; PRIOR FILING DATE: 1999-08-14  
 ; NUMBER OF SEQ ID NOS: 52  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 2  
 ; LENGTH: 328  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-306-686-2

Query Match 100.0%; Score 1731; DB 14; Length 328;  
 Best Local Similarity 100.0%; Pred. No. 4.3e-149; Indels 0; Gaps 0;  
 Matches 328; Conservative 0; Mismatches 0;  
 QY 1 DTFADSLRYVKNILNSKFGFTSRKVPAPMPHMDIRIVMQLQDMFPPEFDKTSFHKYRHS 60  
 DB 1 DTFADSLRYVKNILNSKFGFTSRKVPAPMPHMDIRIVMQLQDMFPPEFDKTSFHKYRHS 60  
 QY 61 EDMQFAFSFYFYLMSAVOPLNISQVFDEVDTDQSGVLSDRIRTLATRIHELPLSLQDLT 120  
 DB 61 EDMQFAFSFYFYLMSAVOPLNISQVFDEVDTDQSGVLSDRIRTLATRIHELPLSLQDLT 120  
 QY 121 GLEHMLNCSKMLPADITQLNNIPTQESYDNPVPVTKSLVTNCKPVTDKIKHAYKOK 180  
 DB 121 GLEHMLNCSKMLPADITQLNNIPTQESYDNPVPVTKSLVTNCKPVTDKIKHAYKOK 180  
 QY 181 NKYFEIMGEETAFKMTNVSHVVGOLDIRKNPKFVCLNDNIDHNHKAQTVKAVL 240  
 DB 181 NKYFEIMGEETAFKMTNVSHVVGOLDIRKNPKFVCLNDNIDHNHKAQTVKAVL 240  
 QY 241 RDFSMPFIPSPQELPREYRNFLHMHQLQEWAYRDKLFWTHCVLATLIMFTIFSFF 300  
 DB 241 RDFSMPFIPSPQELPREYRNFLHMHQLQEWAYRDKLFWTHCVLATLIMFTIFSFF 300  
 QY 301 AEQIALKRRKIFPRRRHKEASPNRIRV 328  
 DB 301 AEQIALKRRKIFPRRRHKEASPNRIRV 328

RESULT 9  
 US-09-895-072-8  
 ; Sequence 8, Application US/09895072  
 ; Patent No. US2002025550A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CANFIELD, WILLIAM M  
 ; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
 ; FILE REFERENCE: 210119SUCCNT

; CURRENT APPLICATION NUMBER: US/09/895,072  
 ; CURRENT FILING DATE: 2001-07-02  
 ; PRIOR APPLICATION NUMBER: 60/153,831  
 ; PRIOR FILING DATE: 1999-09-14  
 ; PRIOR APPLICATION NUMBER: US 09/635,872  
 ; PRIOR FILING DATE: 2000-08-10  
 ; NUMBER OF SEQ ID NOS: 52  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 8  
 ; LENGTH: 328  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 US-09-895-072-8

Query Match 98.1%; Score 1698; DB 9; Length 328;  
 Best Local Similarity 97.6%; Pred. No. 4.4e-146; Indels 0; Gaps 0;  
 Matches 320; Conservative 7; Mismatches 1;  
 QY 1 DTFADSLRYVKNILNSKFGFTSRKVPAPMPHMDIRIVMQLQDMFPPEFDKTSFHKYRHS 60  
 DB 1 DTFADSLRYVKNILNSKFGFTSRKVPAPMPHMDIRIVMQLQDMFPPEFDKTSFHKYRHS 60  
 QY 61 EDMQFAFSFYFYLMSAVOPLNISQVFDEVDTDQSGVLSDRIRTLATRIHELPLSLQDLT 120  
 DB 61 EDMQFAFSFYFYLMSAVOPLNISQVFDEVDTDQSGVLSDRIRTLATRIHELPLSLQDLT 120  
 QY 121 GLEHMLNCSKMLPADITQLNNIPTQESYDNPVPVTKSLVTNCKPVTDKIKHAYKOK 180  
 DB 121 GLEHMLNCSKMLPADITQLNNIPTQESYDNPVPVTKSLVTNCKPVTDKIKHAYKOK 180  
 QY 181 NKYFEIMGEETAFKMTNVSHVVGOLDIRKNPKFVCLNDNIDHNHKAQTVKAVL 240  
 DB 181 NKYFEIMGEETAFKMTNVSHVVGOLDIRKNPKFVCLNDNIDHNHKAQTVKAVL 240  
 QY 241 RDFSMPFIPSPQELPREYRNFLHMHQLQEWAYRDKLFWTHCVLATLIMFTIFSFF 300  
 DB 241 RDFSMPFIPSPQELPREYRNFLHMHQLQEWAYRDKLFWTHCVLATLIMFTIFSFF 300  
 QY 301 AEQIALKRRKIFPRRRHKEASPNRIRV 328  
 DB 301 AEQIALKRRKIFPRRRHKEASPNRIRV 328

RESULT 10  
 US-09-986-552-8  
 ; Sequence 8, Application US/09986552  
 ; Patent No. US20020150981A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CANFIELD, WILLIAM  
 ; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
 ; FILE REFERENCE: 215089US77DIV  
 ; CURRENT APPLICATION NUMBER: US/09/986,552  
 ; CURRENT FILING DATE: 2001-11-09  
 ; PRIOR APPLICATION NUMBER: 09/635,872  
 ; PRIOR FILING DATE: 2000-08-10  
 ; PRIOR APPLICATION NUMBER: 60/153,831  
 ; PRIOR FILING DATE: 1999-09-14  
 ; NUMBER OF SEQ ID NOS: 52  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 8  
 ; LENGTH: 328  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 US-09-986-552-8

Query Match 98.1%; Score 1698; DB 9; Length 328;  
 Best Local Similarity 97.6%; Pred. No. 4.4e-146; Indels 0; Gaps 0;  
 Matches 320; Conservative 7; Mismatches 1;  
 QY 1 DTFADSLRYVKNILNSKFGFTSRKVPAPMPHMDIRIVMQLQDMFPPEFDKTSFHKYRHS 60  
 DB 1 DTFADSLRYVKNILNSKFGFTSRKVPAPMPHMDIRIVMQLQDMFPPEFDKTSFHKYRHS 60



QY 61 EDMQAFSYFYILMSAVQPLNISQVFEVDVTDQSGVLSDRREIRTLATRIHELPLSLQDLT 120  
DB 61 EDMQAFSYFYILMSAVQPLNISQVFEVDVTDQSGVLSDRREIRTLATRIHDLPLSLQDLT 120  
QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYYDPNLPVPTKSLVTNCKPVTDKIKHAYKDK 180  
DB 121 GLEHMLINCSKMLPANITQLNNIPPTQESYYDPNLPVPTKSLVTNCKPVTDKIKHAYKDK 180  
QY 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNIDHNHDKDAQTVCVAVL 240  
DB 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNIDHNHDKDAQTVCVAVL 240  
QY 241 RDFSFMFPPIPSQFELPREYRNRLHMHLEQWEAYRDKLFKFWTHCVLATLIMFTIFSFF 300  
DB 241 RDFSFMFPPIPSQFELPREYRNRLHMHLEQWEAYRDKLFKFWTHCVLATLIMFTIFSFF 300  
QY 301 AEQIALKPKIFPRRRHKEASPNRIRV 328  
DB 301 AEQIALKPKIFPRRRHKEASPNRIRV 328

## RESULT 11

US-10-023-888-10  
; Sequence 10, Application US/10023888  
; Publication No. US20030119088A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: SOLUBLE GLCNAC PHOSPHOTRANSFERASE  
; FILE REFERENCE: 203510US77  
; CURRENT APPLICATION NUMBER: US/10/023,888  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 10  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-10-023-888-10

Query Match 98.1%; Score 1698; DB 14; Length 328;  
Best Local Similarity 97.6%; Pred. No. 4.4e-146;  
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPHPMHMIDRIVMQELQDMFPEEDKTSFHKVVRHS 60  
DB 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPHPMHMIDRIVMQELQDMFPEEDKTSFHKVVRHS 60  
QY 61 EDMQAFSYFYILMSAVQPLNISQVFEVDVTDQSGVLSDRREIRTLATRIHELPLSLQDLT 120  
DB 61 EDMQAFSYFYILMSAVQPLNISQVFEVDVTDQSGVLSDRREIRTLATRIHDLPLSLQDLT 120  
QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYYDPNLPVPTKSLVTNCKPVTDKIKHAYKDK 180  
DB 121 GLEHMLINCSKMLPANITQLNNIPPTQESYYDPNLPVPTKSLVTNCKPVTDKIKHAYKDK 180  
QY 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNIDHNHDKDAQTVCVAVL 240  
DB 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNIDHNHDKDAQTVCVAVL 240  
QY 241 RDFSFMFPPIPSQFELPREYRNRLHMHLEQWEAYRDKLFKFWTHCVLATLIMFTIFSFF 300  
DB 241 RDFSFMFPPIPSQFELPREYRNRLHMHLEQWEAYRDKLFKFWTHCVLATLIMFTIFSFF 300  
QY 301 AEQIALKPKIFPRRRHKEASPNRIRV 328  
DB 301 AEQIALKPKIFPRRRHKEASPNRIRV 328

## RESULT 12

US-10-023-889-10  
; Sequence 10, Application US/10023889  
; Publication No. US20030124652A1  
; GENERAL INFORMATION:

; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS OF PRODUCING HIGH MANNOSE GLYCOPROTEINS IN COMPLEX CARBOH  
; FILE REFERENCE: 203512US77  
; CURRENT APPLICATION NUMBER: US/10/023,889  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 10  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-10-023-889-10

Query Match 98.1%; Score 1698; DB 14; Length 328;  
Best Local Similarity 97.6%; Pred. No. 4.4e-146;  
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;  
QY 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPHPMHMIDRIVMQELQDMFPEEDKTSFHKVVRHS 60  
DB 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPHPMHMIDRIVMQELQDMFPEEDKTSFHKVVRHS 60  
QY 61 EDMQAFSYFYILMSAVQPLNISQVFEVDVTDQSGVLSDRREIRTLATRIHELPLSLQDLT 120  
DB 61 EDMQAFSYFYILMSAVQPLNISQVFEVDVTDQSGVLSDRREIRTLATRIHDLPLSLQDLT 120  
QY 121 GLEHMLINCSKMLPADITQLNNIPPTQESYYDPNLPVPTKSLVTNCKPVTDKIKHAYKDK 180  
DB 121 GLEHMLINCSKMLPANITQLNNIPPTQESYYDPNLPVPTKSLVTNCKPVTDKIKHAYKDK 180  
QY 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNIDHNHDKDAQTVCVAVL 240  
DB 181 NKYRFEIMGEEBIAFKMIRTNVSHVVGQLDDIRKNPKRFVCLNDNIDHNHDKDAQTVCVAVL 240  
QY 241 RDFSFMFPPIPSQFELPREYRNRLHMHLEQWEAYRDKLFKFWTHCVLATLIMFTIFSFF 300  
DB 241 RDFSFMFPPIPSQFELPREYRNRLHMHLEQWEAYRDKLFKFWTHCVLATLIMFTIFSFF 300  
QY 301 AEQIALKPKIFPRRRHKEASPNRIRV 328  
DB 301 AEQIALKPKIFPRRRHKEASPNRIRV 328

## RESULT 13

US-10-023-890-10  
; Sequence 10, Application US/10023890  
; Publication No. US20030124653A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHOD OF PRODUCING GLYCOPROTEINS HAVING REDUCED COMPLEX CARBOH  
; FILE REFERENCE: 203510US77  
; CURRENT APPLICATION NUMBER: US/10/023,890  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 10  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-10-023-890-10

Query Match 98.1%; Score 1698; DB 14; Length 328;  
Best Local Similarity 97.6%; Pred. No. 4.4e-146;  
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;  
QY 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPHPMHMIDRIVMQELQDMFPEEDKTSFHKVVRHS 60  
DB 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPHPMHMIDRIVMQELQDMFPEEDKTSFHKVVRHS 60  
QY 61 EDMQAFSYFYILMSAVQPLNISQVFEVDVTDQSGVLSDRREIRTLATRIHELPLSLQDLT 120  
DB 61 EDMQAFSYFYILMSAVQPLNISQVFEVDVTDQSGVLSDRREIRTLATRIHDLPLSLQDLT 120

QY 121 GLEHMLNCSKMLPADITQLNNIPPTQESYDDNLPVTKSLVTNCKPVTDKIHKAYKDK 180  
 Db 121 GLEHMLNCSKMLPANITQLNNIPPTQESYDDNLPVTKSLVTNCKPVTDKIHKAYKDK 180  
 QY 181 NKYRFEIMGEEIEAFKIRTNVSHVVGQDDIRKNPKFVCLNDNDHNDHDKAQTAVL 240  
 Db 181 NKYRFEIMGEEIEAFKIRTNVSHVVGQDDIRKNPKFVCLNDNDHNDHDKAQTAVL 240  
 QY 241 RDPYESMFPPIPSQPELPREYNRFLHMHQLQEWAYEDKLFVTHCVLATLIMFTIFSFF 300  
 Db 241 RDPYESMFPPIPSQPELPREYNRFLHMHQLQEWAYEDKLFVTHCVLATLIMFTIFSFF 300  
 QY 301 AEQIIALKRKIFPRRRHKEASPNRIRV 328  
 Db 301 AEQIIALKRKIFPRRRHKEASPNRIRV 328

## RESULT 14

US-10-024-197-10  
 ; Sequence 10, Application US/10024197  
 ; Publication No. US20030133924A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: KORNFIELD, William  
 ; TITLE OF INVENTION: HIGHLY PHOSPHORYLATED ACID BETA-GLUCOCEREBOSEDASE AND METHODS  
 ; FILE REFERENCE: 209794USO  
 ; CURRENT APPLICATION NUMBER: US/10/024,197  
 ; CURRENT FILING DATE: 2001-12-21  
 ; NUMBER OF SEQ ID NOS: 27  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 10  
 ; LENGTH: 328  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 US-10-024-197-10

Query Match 98.1%; Score 1698; DB 14; Length 328;  
 Best Local Similarity 97.6%; Pred. No. 4.4e-146;  
 Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPMHMIDRIVMQELQDMFPEEFDKTSFHKVRS 60  
 Db 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPMHMIDRIVMQELQDMFPEEFDKTSFHKVRS 60  
 QY 61 EDMQFAFSFYFLMSAVQPLNISQVFHEVDTQSGVLSDRIRTLATRIHELPLSLQDLT 120  
 Db 61 EDMQFAFSFYFLMSAVQPLNISQVFHEVDTQSGVLSDRIRTLATRIHELPLSLQDLT 120  
 QY 121 GLEHMLNCSKMLPADITQLNNIPPTQESYDDNLPVTKSLVTNCKPVTDKIHKAYKDK 180  
 Db 121 GLEHMLNCSKMLPANITQLNNIPPTQESYDDNLPVTKSLVTNCKPVTDKIHKAYKDK 180  
 QY 181 NKYRFEIMGEEIEAFKIRTNVSHVVGQDDIRKNPKFVCLNDNDHNDHDKAQTAVL 240  
 Db 181 NKYRFEIMGEEIEAFKIRTNVSHVVGQDDIRKNPKFVCLNDNDHNDHDKAQTAVL 240  
 QY 241 RDPYESMFPPIPSQPELPREYNRFLHMHQLQEWAYEDKLFVTHCVLATLIMFTIFSFF 300  
 Db 241 RDPYESMFPPIPSQPELPREYNRFLHMHQLQEWAYEDKLFVTHCVLATLIMFTIFSFF 300  
 QY 301 AEQIIALKRKIFPRRRHKEASPNRIRV 328  
 Db 301 AEQIIALKRKIFPRRRHKEASPNRIRV 328

## RESULT 15

US-10-023-894-10  
 ; Sequence 10, Application US/10023894  
 ; Publication No. US20030143669A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: KORNFIELD, William

; TITLE OF INVENTION: EXPRESSION OF LYSOSOMAL HYDROLASE IN CELLS EXPRESSING PRO-N-  
 ; FILE REFERENCE: 217139US77  
 ; CURRENT APPLICATION NUMBER: US/10/023,894  
 ; CURRENT FILING DATE: 2001-12-21  
 ; NUMBER OF SEQ ID NOS: 22  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 10  
 ; LENGTH: 328  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 US-10-023-894-10

Query Match 98.1%; Score 1698; DB 14; Length 328;  
 Best Local Similarity 97.6%; Pred. No. 4.4e-146;  
 Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPMHMIDRIVMQELQDMFPEEFDKTSFHKVRS 60  
 Db 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPMHMIDRIVMQELQDMFPEEFDKTSFHKVRS 60  
 QY 61 EDMQFAFSFYFLMSAVQPLNISQVFHEVDTQSGVLSDRIRTLATRIHELPLSLQDLT 120  
 Db 61 EDMQFAFSFYFLMSAVQPLNISQVFHEVDTQSGVLSDRIRTLATRIHELPLSLQDLT 120  
 QY 121 GLEHMLNCSKMLPADITQLNNIPPTQESYDDNLPVTKSLVTNCKPVTDKIHKAYKDK 180  
 Db 121 GLEHMLNCSKMLPANITQLNNIPPTQESYDDNLPVTKSLVTNCKPVTDKIHKAYKDK 180  
 QY 181 NKYRFEIMGEEIEAFKIRTNVSHVVGQDDIRKNPKFVCLNDNDHNDHDKAQTAVL 240  
 Db 181 NKYRFEIMGEEIEAFKIRTNVSHVVGQDDIRKNPKFVCLNDNDHNDHDKAQTAVL 240  
 QY 241 RDPYESMFPPIPSQPELPREYNRFLHMHQLQEWAYEDKLFVTHCVLATLIMFTIFSFF 300  
 Db 241 RDPYESMFPPIPSQPELPREYNRFLHMHQLQEWAYEDKLFVTHCVLATLIMFTIFSFF 300  
 QY 301 AEQIIALKRKIFPRRRHKEASPNRIRV 328  
 Db 301 AEQIIALKRKIFPRRRHKEASPNRIRV 328

Search completed: November 21, 2004, 13:22:35  
 Job time : 36.4053 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2004, 12:39:03 ; Search time 10.0204 Seconds  
(without alignments)  
3149.495 Million cell updates/sec

Title: US-10-023-888-5  
Perfect score: 1731  
Sequence: 1 DTFADSLRYNKLNSKPGF.....RKIPFRRRIHKEASPNRIRV 328

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: Pir1.\*  
2: Pir2.\*  
3: Pir3.\*  
4: Pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1290	74.5	248	2 T50618	hypothetical prote
2	581	33.6	384	2 I49528	hypothetical prote
3	124	7.2	1312	1 BMYDL	RAD50 protein - ye
4	113.5	6.6	545	2 C82014	probable capsule b
5	111	6.4	3119	2 T18414	protein g377 - mal
6	106.5	6.2	971	2 T00394	GTP-binding protei
7	104.5	6.0	1999	1 S21801	myosin heavy chain
8	103	6.0	690	2 P71239	hypothetical prote
9	102.5	5.9	706	2 D90124	hypothetical prote
10	102	5.9	515	2 S59811	vacuolar segregati
11	101.5	5.9	301	2 AHI144	probable membrane
12	101	5.8	2663	1 S28261	centromere protein
13	99.5	5.7	587	2 S20596	immediate early pr
14	99	5.7	1961	1 A61231	myosin heavy chain
15	97.5	5.6	735	2 D70174	methy/-accepting c
16	97.5	5.6	1498	2 AG1439	B. subtilis yuka p
17	97	5.6	586	2 T20009	hypothetical prote
18	96.5	5.6	574	2 T33794	hypothetical prote
19	96	5.5	586	2 T34703	hypothetical prote
20	95.5	5.5	1837	2 T41023	probable nuclear p
21	95	5.5	651	2 P89798	hypothetical prote
22	95	5.5	1252	2 B42771	reticulocyte-bindl
23	94	5.4	718	2 G97243	transcription acce
24	94	5.4	810	2 A49744	AMP deaminase (EC
25	94	5.4	817	2 S53921	hypothetical prote
26	94	5.4	1979	1 S03166	myosin heavy chain
27	93.5	5.4	353	2 E70100	tryptophan-tRNA li
28	93.5	5.4	1101	2 AB2481	hypothetical prote
29	93	5.4	314	2 A71614	syntaxin PFE0460W

conserved hypothet  
hypothetical prote  
apolipoprotein A-1  
hypothetical prote  
lantibiotic mersac  
probable multiple  
hypothetical prote  
H+-exporting ATPas  
unknown protein. 7  
DNA-directed DNA p  
DNA-directed DNA p  
hypothetical prote  
peptide synthetase  
dTDP-glucose 4-6-d  
trans-activating t  
early gene transac

## ALIGNMENTS

### RESULT 1

T50618  
hypothetical protein DKFZp762B226.1 - human (fragment)

C;Species: Homo sapiens (man)  
C;Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 09-Jul-2004  
C;Accession: T50618  
R;Bloecker, H.; Boecher, M.; Brandt, P.; Mewes, H.W.; Weil, B.; Wiemann, S.  
submitted to the Protein Sequence Database, June 2000  
A;Reference number: 225143

A;Accession: T50618  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-248 <AAA>  
A;Cross-references: UNIPROT:Q9NPW9; EMBL:AL359588  
A;Experimental source: adult melanoma (MeWo cell line); clone DKFZp762B226  
C;Genetics:  
A;Note: DKFZp762B226.1

Query Match 74.5%; Score 1290; DB 2; Length 248;  
Best Local Similarity 100.0%; Pred. No. 7.6e-87;  
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 VFDEVDTDGSLVSDREIRIATRIHEIPLSIDLTGLEHMLNCKMLPADITQLNNIP 144  
Db 5 VFDEVDTDGSLVSDREIRIATRIHEIPLSIDLTGLEHMLNCKMLPADITQLNNIP 64  
QY 145 PTQESYDNLPPVTKSLVTNCKPVTDKIKHAYKDKNKYRFEIMGEEELAFKMRTNVSH 204  
Db 65 PTQESYDNLPPVTKSLVTNCKPVTDKIKHAYKDKNKYRFEIMGEEELAFKMRTNVSH 124  
QY 205 VVGQLDDIRKNPRKFCVCLNDNDHDKDAQTVAVLRFYFESMPPIPSQFELPREYRNF 264  
Db 125 VVGQLDDIRKNPRKFCVCLNDNDHDKDAQTVAVLRFYFESMPPIPSQFELPREYRNF 184  
QY 265 LHMHELOEQRAYRDKLKFVTHCVLATLINTFTFSFAEQLIALKRIPRRRIHKEASN 324  
Db 185 LHMHELOEQRAYRDKLKFVTHCVLATLINTFTFSFAEQLIALKRIPRRRIHKEASN 244  
QY 325 RIRV 328  
Db 245 RIRV 248

### RESULT 2

I49528  
hypothetical protein - mouse (fragment)

N;Alternate names: CDNA1  
C;Species: Mus musculus (house mouse)  
C;Date: 09-Mar-1996 #sequence\_revision 09-Mar-1996 #text\_change 09-Jul-2004  
C;Accession: I49528  
R;Cordes, S.P.; Barsh, G.S.  
Cell 79, 1025-1034, 1994

A:Title: The mouse segmentation gene kr encodes a novel basic domain-leucine zipper transmembrane protein  
A:Reference number: A55200; MUID:95094266; PMID:8001130

A:Accession: T49528

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-384 <RES>

A:Cross-references: UNIPROT:Q61340; GB:L36434; NID:G625041; PID:AAA65688.1; PID:G625042

Query Match 33.6%; Score 581; DB 2; Length 384;

Best Local Similarity 97.4%; Pred. No. 5.9e-35;

Matches 112; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DTFADSLRYNKNILNSKFGTSSKVPAPMPHMDIRVMQELQDMFPPEEDKTSFHKVRHS 60

Db 267 DTFADSLRYNKNILNSKFGTSSKVPAPMPHMDIRVMQELQDMFPPEEDKTSFHKVRHS 326

Qy 61 EDMQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDEIRTLATRIHELPLS 115

Db 327 EDMQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDEIRTLATRIHDPLT 381

Qy 61 EDMQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDEIRTLATRIHELPLS 115

Db 327 EDMQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDEIRTLATRIHDPLT 381

RESULT 3

BWBYDL

N:Alternate names: protein N0872; protein YNL250W

C:Species: Saccharomyces cerevisiae

C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 09-Jul-2004

C:Accession: S05808; S63223

R:Alani, E.; Subbiah, S.; Kleckner, N.

Genetics 122, 47-57, 1989

A:Title: The yeast RAD50 gene encodes a predicted 153-kD protein containing a purine nucleotide-binding domain

A:Reference number: S05808; MUID:8927691; PMID:2659437

A:Accession: S05808

A:Molecule type: DNA

A:Residues: 1-1312 <ALA>

A:Cross-references: UNIPROT:P12753; EMBL:X14814; NID:G4272; PID:CAA32919.1; PID:G4273

R:Sen-Gupta, M.; Guelider, U.; Beinbauer, J.; Fiedler, T.; Hegemann, J.H.

submitted to the Protein Sequence Database, April 1996

A:Reference number: S63220

A:Accession: S63223

A:Molecule type: DNA

A:Residues: 1-1312 <SEN>

A:Cross-references: EMBL:Z71526; NID:G1302292; PIDN:CAA96157.1; PID:G1302293; GSPDB:GN00

A:Experimental source: strain S288C

C:Genetics: S63220

A:Gene: SGD:RAD50; MIPS:YNL250W

A:Cross-references: SGD:S0005194; MIPS:YNL250W

A:Map position: 14L

C:Superfamily: RAD50 protein

C:Keywords: ATP; coiled coil; DNA repair; meiosis; nucleus

F:177-421/Region: heptad repeats

F:743-995/Region: heptad repeats

F:407/Binding site: ATP (lys) #status predicted

Query Match 7.2%; Score 124; DB 1; Length 1312;

Best Local Similarity 21.3%; Pred. No. 0.54;

Matches 68; Conservative 65; Mismatches 126; Indels 60; Gaps 14;

Qy 5 DLSRYNKNILNSKFGTSSKVPAPMPHMDIRVMQELQDMFPPEEDKTSF 54

Db 189 DNLKSIKXDMVDIKLQKQSV-EHLKLDKDSKAWKLNHQLQYKIDQYNEVSEIESQL 247

Qy 55 HKVRHSDMQAFSY-FYLLMSAVQPLNISQVFEVDVTDQSGVLSDEIRTLATRIHELPL 113

Db 248 NEITEKSKDKFKNQDFQKILSKVENLKVTKL-----SISD-QVKRLSNSIDLD 296

Qy 114 LSLQDLTGLEHMLNCSKMLPADITQLNIPPTQESYVD--PNLPVTVKSLVNTCKPVT 171

Db 297 LSKPD---LQNLNANFSKVLMDKXNKQSLDLETDISLKDQSSQLSNLIRKQGL-- 351

Qy 172 KHKAYKDKNYFRIMGEIEAFKQVIRTVSHV-----VGQLDDIRKNPKFVC--LNDN 225

Db 352 EAGKETYEKRNHLSLKE---AFQHKFQGLSNIENSMDAQVNHMSQKRAFIQDLTDT 408

Qy 226 IDHNHKAQTVKAVLRDYPFYSMPPIPQSFELPREYRNFLLH----- 266

Db 409 IDQFAKDITQLKETNLSDLIKSITVDQSNLEYNKDRSKLIHDSLEAEKLUKSPKSLSTQD 468

Qy 267 --MHELOEWRAIRDKLKFV 283

Db 469 SLNHELENLKYKELQSW 487

RESULT 4

CS2014

A:Title: probable capsule biosynthesis protein NMA0200 [imported] - Neisseria meningitidis (strain

C:Species: Neisseria meningitidis

C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 09-Jul-2004

C:Accession: CS2014

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel

; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream

Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.

A:Reference number: A81775; MUID:2022556; PMID:10761919

A:Accession: CS2014

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-545 <PAR>

A:Cross-references: UNIPROT:Q9JWW8; GB:AL162752; GB:AL157959; NID:G7378778; PIDN:CA8835

A:Experimental source: serogroup A, strain Z2491

C:Genetics:

A:Gene: sacB; NMA0200

Query Match 6.6%; Score 113.5; DB 2; Length 545;

Best Local Similarity 18.4%; Pred. No. 1;

Matches 47; Conservative 36; Mismatches 65; Indels 107; Gaps 8;

Qy 3 FADSLRYNKNILNSKFGTSSKVPAPMPHMDIRVMQELQDMFPPEEDKTSFHKVRHSD 62

Db 395 YLNGARNANTLLEKFKFTKJHSPQSWRTDILFEMEXKYFEENRNLHNFSLDD 454

Qy 63 MQF-AFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDEIRTLATRIHELPLSLQDLTG 121

Db 455 IAVTGYLYHHV-----ALLSGRALQS-----SDKTE 480

Qy 122 LEHMLNCSKMLPADITQLNIPPTQESYVDYDNLPPVTKSLVNTCKPVTDKIHKAYKDK 180

Db 481 LVQQNHDFKK-----KLNWVTLTKERNFD-KLP----- 508

Qy 181 NKRYFEIMGEBEIAFPMIRTVSHVVGQLDDIRKNPKRFVCLNDINHHKDAQTVKAVL 240

Db 509 -----LSVCINDGADSHLNEWNVQVI- 530

Qy 241 RDPFYSMPPIPQSF 255

Db 531 -KFELETLPPLPSSFE 544

Qy 241 RDPFYSMPPIPQSF 255

Db 531 -KFELETLPPLPSSFE 544

Qy 241 RDPFYSMPPIPQSF 255

Db 531 -KFELETLPPLPSSFE 544

Qy 241 RDPFYSMPPIPQSF 255

Db 531 -KFELETLPPLPSSFE 544

RESULT 5

T18414

A:Title: protein g377 - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T18414

R:Handman, E.; Osborn, A.H.; Symons, F.; Van Driel, R.; Cappai, R.

Mol. Biochem. Parasitol. 74, 143-156, 1995

A:Title: The Leishmania promastigote surface antigen 2 complex is differentially expressed

A:Reference number: Z18933; MUID:96360472; PMID:8719156

A:Accession: T18414

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3119 <HAN>

A:Cross-references: UNIPROT:Q25857; EMBL:L04161; NID:G309687; PID:G309688; PIDN:AAAC3725

Query Match 6.4%; Score 111; DB 2; Length 3119;

Best Local Similarity 22.7%; Pred. No. 14;

```

Matches 68; Conservative 50; Mismatches 94; Indels 88; Gaps 16;
QY 41 LQDMFPEDFTSKTSEHKVRHSDMOPAFSYFYFYLMSAVQPLNISQVDFEVDTDQSGVLSD 99
Db 1261 LEDIYME--KKKYKEEYSKMRRISSNLDYEVNK---QIKEHYHKVDI-----ISEH 1308
QY 100 --REIR-----TLATRIHELPLSLQ-----DLTGLEHMLINCSKMLPADITQLANNIPP-- 145
Db 1309 KFQELRQHMRKIENTHELKYMVYQIQIDLTYYHQLENIHSELLQALQONKNIPIRLH 1368
QY 146 -----TOESYDYNLPVP--TKSLVTNCKPVTDKIHKAYKD-----KKKYRFEI--- 187
Db 1369 NVLEKKLEITKRK--KKQKPDISTSSHATDEQQVSDTLIRGAHNHGDIKGEDNDEVLLI 1426
QY 188 -----MGEELIAPKMIRTNVSHVVGQ-----LDDIRKPRKPVCLNDNIDH 228
Db 1427 EQISLTKMGDNQNGVSIKLEKLNDSQYQLQDKLVVVEDIYKLNRF---KHVIEK 1483
QY 229 NHKDAQ-----TVKAVLRDFEYSMPFIPSQFELPREYRNRFLEHMELOEWAYRDKLK 281
Db 1484 LHKESKINREKFIITKVDVLSNVYSTL-----EYVVKEL-LHDFQEWSEKDELE 1531

RESULT 5
T00394
GTP-binding protein - cherry salmon
C/Species: Oncorhynchus masou (Cherry salmon)
C/Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 09-Jul-2004
C/Accession: T00394
R/Kubokawa, K.; Miyashita, T.; Kubo, Y.
FEBS Lett. 431, 231-235, 1998
A/Title: Isolation of a cDNA for a novel 120-kDa GTP-binding protein expressed in motor
A/Reference number: Z14145; MUID: 99372640; PMID: 9708909
A/Accession: T00394
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-971 <KUB>
A/Cross-references: UNIPROT:O93248; EMBL:AB012720; NID:G3413519; PIDN:BAJ32279.1; PID:G34
A/Genes: mg120
C/Keywords: GTP binding; nucleotide binding; P-loop
F/307-314/Region: nucleotide-binding motif A (P-loop)

```

	Query Match	6.2%; Score 106.5; DB 2; Length 971;
	Best Local Similarity	23.0%; Pred. No. 6.9;
	Matches 62; Conservative	43; Mismatches 79; Indels 85; Gaps 14;
Qy	7 LRYVNKILNSKF--GFTSRKVPAAHPHMDIRIVMQLQDMPEEFDTSTFKHYRHSDDMQ 64     :   :    :	:
Dd	102 LRYI--VLGSAVGGGYTAK-----TYEWNKMPLD----- 130 :	:  :  :
Qy	65 FAFSYFYLLMSA----VOPLNISVFDEVDTDQ--SGVLSDREIRTATRIHELPLESLQDLT 120 :    :  :  :  :	:  :
Dd	131 -----MSAYNWVIPDFWVELSQIDIDLKTKILP--ELEBIAKLLPELP---DF- 174 :	:
Qy	121 GLEHMLINCSKLMPADITQLNNIPTGESYYDNLPPVTIKSVLNCTKPVTDKHKYVKDK 180     :  :  :  :	:
Dd	175 -----DKIGENFTELKSILFTAEPGDTPKAATEAPVTATEASDKQPKSSDX 224 :	:
Qy	181 NKYEFEINGBEEIAFPKIRTNVSH--VVGLDDLTRKPRXPCVCLNDNID-HNHKDQAOTVK 237 :    :  :  :  :	:  :
Dd	225 EKVG-----DOLEEILLRTQMKYRWLERKEKENKOLRVKVLOKDEKGHIHQRK-----IK 273 :	:
Qy	238 AVLRFDYESMFPIPSQE-----LPR 258 :	:
Dd	274 KSIIDMTSEVIDILDSPDSNYNTODHLPR 302 :	:

RESULT 7  
S21801  
myosin heavy chain, neuronal [similarity] - rat  
N;Alternate names: myosin II  
N;Contains: myosin ATPase [EC 3.6.4.1]

C:Species: Rattus norvegicus (Norway rat)  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 09-Jul-2004  
C:Accession: S21801; #PNU0013; S18134  
R:Sun, W.; Chantler, P.D.  
J. Mol. Biol. 224, 1185-1193, 1992  
A:Title: Cloning of the cDNA encoding a neuronal myosin heavy chain from mammalian brain  
A:Reference number: S21801; MUID:92235856; PMID:1569576  
A:Accession: S21801  
A:Molecule type: mRNA  
A:Residues: 1-1999 <SUN>  
A:Cross-references: UNIPROT:Q63731; EMBL:X62659  
R:Sun, W.; Chantler, P.D.  
Biochem. Biophys. Res. Commun. 175, 244-249, 1991  
A:Title: A unique cellular myosin II exhibiting differential expression in the cerebra  
A:Reference number: PNU013; MUID:91151356; PMID:1998509  
A:Accession: PNU013  
A:Molecule type: mRNA  
A:Residues: 1914-1998, 'I' <SU2>  
A:Experimental source: brain  
C:Superfamily: myosin heavy chain; myosin motor domain homology  
C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleot  
F:84-763/Domain: myosin motor domain homology <MMOT>  
F:174-181/Region: nucleotide-binding motif A (P-loop)  
F:541-575/Region: actin binding #status predicted  
F:653-675/Region: actin binding #status predicted  
F:836-1999/Domain: coiled coil #status predicted <COI>  
F:836-1276/Region: S2  
F:1277-1999/Region: light meromyosin  
F:125/Modified site: N6,N6-trimethyllysine (Lys) #status predicted  
F:180/Binding site: ATP (Lys) #status predicted  
F:693,703/Active site: Cys #status predicted  
F:1916/Binding site: phosphate (Ser) (covalent) #status predicted  
F:1943/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 6.0%; Score 104.5; DB 1; Length 1999;  
Best Local Similarity 26.5%; Pred. No. 24;  
Matches 40; Conservative 29; Mismatches 63; Indels 19; Gaps 5;

Qy 99 DRETRTLATRIHELPLSLQDLTGLEHMLNCSKMLPAITQL-NNIPPTQESYDNLPP 157  
Db 1364 ERRVTEADKVTKLQVELNDVTGLSQSDSKSKLTQPSALESQLDQTQELLQENRQK 1323  
Qy 158 VTKSLVNCNKVDTKIKAYKDKNKRYPEIMGEIEIAFKMIRTNVSHVVGQLDDTRKNPR 217  
Db 1324 L--SLSTKLQVED-----EKNSPRLQLEEEEEKHNLEKQIATLHAQVADMKKQME 1374  
Qy 218 KTV-CLNDNDHNKQATQTVKAVLRDFYEM 247  
Db 1375 DSVGCL-----ETAEEVKRLQKQLEGL 1397

RESULT 8  
F71239  
hypothetical protein PH0176 - Pyrococcus horikoshii  
C:Species: Pyrococcus horikoshii  
C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 09-Jul-2004  
C:Accession: F71239  
R:Kawakabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Seki  
M.; Onfuku, Y.; Funabaehi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi  
DNA Res. 5, 55-76, 1998  
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic  
A:Reference number: A71000; MUID:98344137; PMID:9679194  
A:Accession: F71239  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-690 <KAW>  
A:Cross-references: UNIPROT:O57915; GB:AP000001; NID:G3236128; PIDN:BA029245.1; PID:G3  
A:Experimental source: strain OT3  
A:Note: This accession replaces an interim accession for a sequence replaced by GenBank  
C:Genetics:  
A:Gene: PH0176

Query Match

[illegible]

## RESULT 9

RESULT 9  
D90124  
hypothetical protein rad3 [imported] - Guillardia theta nucleomorph  
C:Species: nucleomorph Guillardia theta  
A:/Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont  
C:/date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:/Accession: D90124  
R:/Doughlas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Reid,  
Nature 410, 1091-1096, 2001  
A;/title: The highly reduced genome of an enslaved algal nucleus.  
A;/reference number: A99082; MUID:11323671; PMID:11323671  
A:/Accession: D90124  
A;/Status: preliminary  
A;/Molecule type: DNA  
A;/Residues: 1-706 <DOU>  
A;/Cross-references: UNIPROT:Q98S94; GB:AF083031; NID:g13794312; PIDN:AAK39689.1; GSPDB:G  
C:/Genetics:  
A;/Gene: rad3  
A;/Map position: 3  
A;/Genome: nucleomorph  
C:/Keywords: nucleomorph

Query Match	5.9%	Score 102.5	DB 2	Length 706
Best Local Similarity	21.1%	Pred. No. 9.1		
Matches	77	Conservative 55	Mismatches 144	Indels 75
Gaps	17			
Qy	6	SLRYVNYKILNSKFGFTSR-KVPAHWPMID-----RIVMQELQDMFPPEFDKTS-----	53	
Db	135	SLDVETSSLSNKSDFVNSKFLKPNCFNVFLLKRRKIYKVYIDTRKILKFSRNNKVCY 194		
Qy	54	FHKVTHSEDMQPAFSFYFLMSAVOPLNISQ-----VEDEVDTQSOGLVSDREI-----	102	
Db	195	FFILANVEDYNFLILKPDQIIHSSKNYLNATKNKKIPLFSPNES-FSALLKSKSIISLS 253		
Qy	103	-RLATRIHELPLSLQDITGIEHMLINCMSKMLPADITQLNNIPTTOESY-DPNLPPVTK 160		
Db	254	LKTIIDSYSLSIIVERDIIKKIKQIVNLK-----TFNNFLKKNFVYKKNKKILNN 306		
Qy	161	SLVTNCKPVDKIHKAYKDKKNKYFEINGEETAFKMRITNVSHVVGOL-----D 210		
Db	307	YLVTKLNFONKFLHSLY-LNSYRFL-----ETLLKLSKKEKYISSINIYKDLFLEN 361		
Qy	211	DIRKQPRKVCVLNDNIDH-----NHKAQTVKAVLRDFYSEMFPIPSQFELPREYN 262		
Db	362	FTEIAPVYFONFNDNI SHWCKTYSFEDNKSUNGKKI-----SNFLIITKFFL-NYQ 413		

QY 263 RFLHM-HELQEWPRAYRDKLKEWTHCVLATLIMFTIFSFAEQIALKKKIF 312

D6 414 NFNINIKSELKNNSEKPKL-----TLIEDNEGIFEEFFEE---LPSIF 454

## RESULT 10

S59811  
 vacuolar segregation protein PEP7 - yeast (Saccharomyces cerevisiae)  
 N;Alternate names: protein D9798.11; protein YDR333c; vacuole segregation protein VAC1  
 C;Species: Saccharomyces cerevisiae  
 C;Date: 13-Jan-1996 #sequence revision 01-Mar-1996 #text\_change 09-Jul-2004  
 C;Accession: S59811; S59789; A41756  
 R;Webb, G.C.  
 submitted to the EMBL Data Library, March 1995  
 A;Reference number: S59811  
 A;Accession: S59811  
 A;Molecule type: DNA  
 A;Residues: 1-515 <WEB>  
 A;Cross-references: UNIPROT:Q99229; EMBL:U22070; NID:G736413; PIDN:AAB60290.1; PID:G736413  
 A;Experimental source: strain S288C  
 R;Du, Z.  
 submitted to the EMBL Data Library, July 1995  
 A;Description: The sequence of S. cerevisiae cosmid 9798.  
 A;Reference number: S59418  
 A;Accession: S59789  
 A;Molecule type: DNA  
 A;Residues: 1-515 <DU>  
 A;Cross-references: EMBL:U22517; NID:G914989; PIDN:AAB64759.1; PID:G915000; MIPS:YDR3233  
 A;Experimental source: strain S288C (AB972)  
 R;Weisman, L.S.; Wickner, W.  
 J. Biol. Chem. 267, 618-623, 1992  
 A;Title: Molecular characterization of VAC1, a gene required for vacuole inheritance and  
 A;Reference number: A41756; MUID:92112720; PMID:1730622  
 A;Accession: A41756  
 A;Molecule type: DNA  
 A;Residues: 1-365,'Y',367-515 <WEI>  
 A;Cross-references: GB:M80596; NID:G173156; PIDN:AAA35203.1; PID:G173157  
 C;Genetics:  
 A;Gene: SGD:PEP7; VAC1; VPS19; VPT19; VPL21  
 A;Cross-references: SGD:S0002731; MIPS:YDR323C  
 A;Map position: 4K  
 C;Function:  
 A;Description: required for vacuole segregation and vacuolar protein sorting; required  
 pathways

```

Query Match      5.9%; Score 102; DB 2; Length 515;
Best Local Similarity 22.8%; Pred. No. 6.6;
Matches 58; Conservative 38; Mismatches 96; Indels 62; Gaps 12

Qy      28 HMP-----HMIDRVMQEIQDMPPEEFKTSFHKVRHSEDMQFAFSYFYILMSAVQPL 80
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      281 HIPISIRLCSHCIDWL-----FGRKFNK-----DVRMPLSGIFAKYDSMQ-- 321

Qy      61 NTSQVFDEVDQDQSVLSDRIPLATRIHELPLSLQDLTGLEHLLNCSKMLPADITQ 139
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      322 NISKVIDSLLPIFEDSLNSLKVEAKDSENTLOPKNLNDLARURHKLLNSFNDYNTLTRQ 381

Qy      140 L-----NNIPPTQESYDYPNPFPVTKSLIVTNCKEVTDKIHKAYDKNKNYRF 185
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      382 LLSVPEQSHLERQLQNSIKIASAYINEKILPLKSLPALINL---EGHKTNEDGQKAP 437

Qy      186 EI-----MGEEIAPK-----WRTNVSHVVQLQDDIHKNPR--KFVCLNDNIDH 228
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      438 EYKQSGQMENLTITKEVKEUREELMWIKQSYLIESTIQDYKKQRRLSEIVTLNKLQEE 497

Qy      229 NHKDAQTKVAKVLRD 242
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      498 LFSRIHTVOSKLG 511
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```



## RESULT 11

AH1144  
Probable membrane protein lmo0559 [imported] - Listeria monocytogenes (strain EGD-e)  
C:Species: Listeria monocytogenes  
C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
C:Accession: AH1144  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duseurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vasquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative Genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11579669  
A:Accession: AH1144  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-301 <GLA>  
A:Cross-references: UNIPROT:Q8Y9G9; GB:NC\_003210; PIDN:CAC98638.1; PID:g16409935; GSPDB:  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo0559

Query Match 5.9%; Score 101.5; DB 2; Length 301;  
Best Local Similarity 20.2%; Pred. No. 3.7;  
Matches 53; Conservative 44; Mismatches 82; Indels 83; Gaps 12;  
QY 57 VRHSDM---QFASYPYILMSAVQPLNISQVFEVDVTDQSGVLSDRSREITLARIHELP 113  
Db 109 LKESDEVTIFOLSSSYFIMDAFPF-----YVEMDTDRITNDLKIKITTKNL----- 159  
QY 114 LSLQDL-TGLEHMLINCSKMLPADITQLNNIPPTQESYVDPLPPVTKSLVTCKPVTDK 172  
Db 160 LSLSDLETGIVY-FVSASKQNAALLEQM-----K 187  
QY 173 IHKAYDKNKYRFIMGEIEIAFKMI--RTNV-SHVVGQLDDIRKNPRKFCVCLNDINHN 229  
Db 188 AHMYRELNEVEKEQFDALIEARQLVEMTGLSSQILQQLSGTYNN-----ILNNLNLD 242  
QY 230 HKDAQ-----TKVALRDFVESMFPIPSQFELPREYRNFLHMHQEWRAVDKLFK 283  
Db 243 MKILTAISILTVTITGTGFGNMPLPLE-----HNTF-----GW 278  
QY 284 THCVLATLIMTIFSFFAEQLI 305  
Db 279 IVTIFISVILWFLGSLFKLKM 300

## RESULT 12

S28261  
centromere protein E - human  
N:Alternate names: centromere 312K protein; kinesin-related protein CENP-E  
C:Species: Homo sapiens (man)  
C>Date: 03-Mar-1994 #sequence\_revision 03-Mar-1994 #text\_change 09-Jul-2004  
C:Accession: S28261  
R:Yen, T.J.; Li, G.; Schaar, B.T.; Szilak, I.; Cleveland, D.W.  
Nature 359, 536-539, 1992  
A:Title: CENP-E is a putative kinetochore motor that accumulates just before mitosis.  
A:Reference number: S28261; MUID:93024922; PMID:1406971  
A:Accession: S28261  
A:Molecule type: mRNA  
A:Residues: 1-2663 <YEN>  
A:Cross-references: UNIPROT:Q02224; EMBL:Z15005; NID:929864; PIDN:CAA78727.1; PID:g29865  
C:Genetics:  
A:Gene: GDB:CENPE  
A:Cross-references: GDB:361164; OMIM:117143  
A:Map position: 4q24-4q25  
C:Superfamily: centromere protein E; kinesin motor domain homology  
C:Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop  
F:7-335/Domain: kinesin motor domain homology <KMOF>  
F:86-93/Region: nucleotide-binding motif A (P-loop)  
F:486-2183/Domain: coiled coil #status predicted <COI>

## F:92/Binding site: ATP (Lys) #status predicted

Query Match 5.8%; Score 101; DB 1; Length 2663;  
Best Local Similarity 20.5%; Pred. No. 62;  
Matches 60; Conservative 47; Mismatches 127; Indels 58; Gaps 10;  
QY 33 IDRIVVCQELQDFPEPFEDKTSFHKVHSDMQFASFYFYLMASAVQPLNISQVFEVD 92  
Db 1972 IQELQKQLQLRVDEVDNMHKKINEMEQKKQKQEPNYLCKCEMDNFQLTKKUHE---- 2027  
QY 93 QSGVLSDRSREITLARIHELPISLQDLTGLEHMLINCSKMLPADITQLNNIPPTQESYVD 152  
Db 2028 ----SLESIRIVAXERDELRIKESLQWERDQFIATLREMIARDQNHQVKEPKRLSD 2082  
QY 153 PNLPPVTKSLVTNCKPVTDKIHKAYDKNKY-----REIMGEIEIAFKMRTNVSHVG 207  
Db 2083 -QQHMLSELRKCSRIKELL-KRYSEMDHDYECNLRLSLDLLEKIEFHRIMKKLYLS 2140  
QY 208 QLLDDIRKNPRKFCVCLND-----NIDNHNKDAQT----- 235  
Db 2141 YVTKIEEQHE--CINKFEMDFIDEVKKELLIKIQLHQDCDVPSELRLDLKLQNWD 2198  
QY 236 --VKAVLRDFYBSMFP-IPSQFELPREYRNFLHMHQEWRAVR---DKLK 281  
Db 2199 LHIEBILKDFSEFPFSIKTEFQ--QVLSNRKEMTQFLEWLNTRFDIEKLK 2248  
RESULT 13  
S20596  
immediate early protein - Bombyx mori nuclear polyhedrosis virus  
C:Species: Bombyx mori nuclear polyhedrosis virus, BmSNPV  
C>Date: 22-Nov-1993 #sequence\_revision 01-Dec-1995 #text\_change 09-Jul-2004  
C:Accession: S20596  
R:Ruybrechts, R.; Guarino, L.; van Brussel, M.; Vulsteke, V.  
Biochim. Biophys. Acta 1129, 328-330, 1992  
A:Title: Nucleotide sequence of a transactivating Bombyx mori nuclear polyhedrosis virus  
A:Reference number: S20596; MUID:92162753; PMID:1536885  
A:Accession: S20596  
A:Status: preliminary; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-587 <HUY>  
A:Cross-references: UNIPROT:P33245; EMBL:X58442; NID:958714; PIDN:CAA41348.1; PID:g5871  
C:Superfamily: Autographa californica nuclear polyhedrosis virus trans-activating trans  
C:Keywords: transcription regulation

Query Match 5.7%; Score 99.5; DB 2; Length 587;  
Best Local Similarity 23.4%; Pred. No. 12;  
Matches 58; Conservative 34; Mismatches 103; Indels 53; Gaps 11;  
QY 39 QELQDMFPBE-----FDKTSFHKVRHSDM-QFASFYFYLMASAVQPLNI-----SQ 84  
Db 189 QBITHYFTNDFAPYLMRFDDNDYNSRFSMDHSETGYTMFVYKSEVKPFPIIFAKYVSN 248  
QY 85 VPDEVDTQSGVLSDRSREITLARIH-----ELPLSLQDLTGLEHMLINCSK 131  
Db 249 VVVEY-TNNYVMDNRVFWTFDKRFMLSYNLVKETGIEIPHS-QDVCDETAQNCKK 306  
QY 132 MLPADITQLNNIPPTQESYVDPLPPVTKSLVT-----NCKPVTDKIHKAYDKNKY 183  
Db 307 CHFVDVH--HTFKAALTSYVNDMYAQTTFVTLQSLGERKCGFLSKLYEMVQDKNLF 364  
QY 184 REEIM-----GEEIIAKF--MIRNVSHVVGQLDDIRK-----NPRKFCVCLNDINHN 229  
Db 365 TLPIMLSRKESNEIETAGNNFVSPYVSQILKYSIRKVPFPDPNPKYVVDNLNLI 424  
QY 230 HXDAQTVK 237  
Db 425 KISTLTLYK 432

## RESULT 14

A61231  
myosin heavy chain nonmuscle form A - human



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2004, 12:46:58 ; Search time 47.0289 Seconds  
(without alignment)  
4012.911 Million cell updates/sec

Title: US-10-023-888-5

Perfect score: 1731

Sequence: 1 DTFADSLRYVNKILNSKFGF.....RKIPRRRIHKEASPNRIRV 328

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_02.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1731	100.0	490	2 Q86TQ2	Q86TQ2 homo sapien
2	1731	100.0	950	2 Q9ULJ2	Q9ULJ2 homo sapien
3	1698	98.1	921	2 Q6P9R5	Q6P9R5 mus musculus
4	1698	98.1	921	2 AAH60638	AAH60638 mus muscu
5	1290	74.5	248	2 Q9NPW9	Q9NPW9 homo sapien
6	704	40.7	132	2 Q9NV34	Q9NV34 homo sapien
7	581	33.6	384	2 Q61340	Q61340 mus musculus
8	553	31.9	717	2 Q7Q098	Q7Q098 anopheles g
9	506	29.2	666	2 Q9V553	Q9V553 drosophila
10	506	29.2	666	2 Q8SX14	Q8SX14 drosophila
11	124	7.2	1312	1 RA50_YEAST	P12753 saccharomyc
12	117.5	6.8	691	2 Q6FKK0	Q6FKK0 candida gla
13	113.5	6.6	384	2 Q6JUT9	Q6JUT9 pichia angu
14	113.5	6.6	384	2 AAQ74772	AAQ74772 pichia an
15	113.5	6.6	545	2 Q68215	Q68215 neisseria m
16	113.5	6.6	545	2 Q83U59	Q83U59 neisseria m
17	113.5	6.6	545	2 Q84CZ9	Q84CZ9 neisseria m
18	113.5	6.6	545	2 Q84D00	Q84D00 neisseria m
19	113.5	6.6	545	2 Q9JWM8	Q9JWM8 neisseria m
20	111	6.4	1048	2 Q26023	Q26023 plasmodium
21	111	6.4	3119	2 Q25857	Q25857 plasmodium
22	111	6.4	3119	2 Q81473	Q81473 plasmodium
23	110	6.4	1072	2 Q6FTD7	Q6FTD7 candida gla
24	108.5	6.3	480	2 Q7R8R2	Q7R8R2 plasmodium
25	106.5	6.2	483	2 Q86KM0	Q86KM0 dictyosteli
26	106.5	6.2	971	1 OPAL_ONCMA	Q93248 oncorhynch
27	106	6.1	574	1 EZRA_STRPY	Q9A0K9 streptococc
28	106	6.1	739	2 Q9M479	Q9M479 physarum po
29	106	6.1	2148	2 Q9BDJ3	Q9BDJ3 physarum po
30	105	6.1	666	2 Q90XS4	Q90XS4 xenopus lae
31	104.5	6.0	301	2 Q722Z1	Q722Z1 listeria mo

32 104.5 6.0 301 2 AAT03370 Aat03370 listeria  
33 104.5 6.0 1999 2 Q63731 Q63731 rattus norv  
34 104 6.0 424 2 Q81I88 Q81I88 plasmodium  
35 104 6.0 574 1 EZRA\_STRP3 Q8k839 streptococ  
36 103.5 6.0 1110 2 Q7RRC2 Q7rrc2 plasmodium  
37 103 6.0 621 2 Q813R0 Q813r0 bacillus ce  
38 103 6.0 690 2 Q57915 Q57915 pyrococcus  
39 102.5 5.9 653 2 Q6BFN4 Q6bfN4 paramecium  
40 102.5 5.9 676 2 Q7YXZ3 Q7yxz3 cryptospori  
41 102.5 5.9 706 2 Q98S94 Q98s94 guillardia  
42 102.5 5.9 1665 2 Q6VA77 Q6va77 plasmodium  
43 102.5 5.9 1665 2 AAO38041 Aao38041 plasmodu  
44 102 5.9 515 2 Q99229 Q99229 saccharomyc  
45 101.5 5.9 301 2 Q8Y9G9 Q8y999 listeria mo

#### ALIGNMENTS

##### RESULT 1

Q86TQ2 PRELIMINARY; PRT; 490 AA.  
AC Q86TQ2;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Ssimilar to v-maf musculoaponeurotic fibrosarcoma oncogene family,  
DE Protein B (Avian) (Fragment).  
OS Homo sapiens (Human).  
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
CN NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC Straussberg R.;  
RA TISSUE=Liver;  
RL Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; BC042615; AAH42615.1; -  
DR GO; GO:0005509; F:calcium ion binding; IEA.  
DR InterPro; IPR010506; DMAP binding.  
DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR010983; EF-hand.  
DR Pfam; PF06464; DMAP\_binding; 1.  
DR Prosite; PS00036; ehand; 1.  
DR PROSITE; PS00018; EF\_HAND; UNKNOWN\_1.  
FT NON TER 1  
SQ SEQUENCE 490 AA; 57411 MW; 37B19FE0D1259AD2 CRC64;

Query Match 100.0%; Score 1731; DB 2; Length 490;  
Best Local Similarity 100.0%; Pred.No. 2e-120;  
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVNKILNSKFGFTRSRKVPAPHPMHIDRIVNQELQDMFPEFEDKTSFHKVRHS 60  
DB 163 DTFADSLRYVNKILNSKFGFTRSRKVPAPHPMHIDRIVNQELQDMFPEFEDKTSFHKVRHS 222  
QY 61 EDMQAFSYFYILMSAVQPLNISQVDFEVDTDQSGVLSDRREIRTLATRIHELPLSLQDLT 120  
DB 223 EDMQAFSYFYILMSAVQPLNISQVDFEVDTDQSGVLSDRREIRTLATRIHELPLSLQDLT 282  
QY 121 GUEHMLINCSKMLPADITQLANNIPPTQSYVDNLPVTKSLVTNCKPVTDKHKAYKDK 180  
DB 283 GUEHMLINCSKMLPADITQLANNIPPTQSYVDNLPVTKSLVTNCKPVTDKHKAYKDK 342  
QY 181 NKYRFEIINGEETIAFMKIRTNVSHVVGQLDIDIRKNPKRFVCLINDNIDHNHDKDAQTVKAVL 240  
DB 343 NKYRFEIINGEETIAFMKIRTNVSHVVGQLDIDIRKNPKRFVCLINDNIDHNHDKDAQTVKAVL 402  
QY 241 RDFSMSPIPSQFELPREYNRFHMHLEQSWRAYRDKLFWTHCVLATIMFTIFSFF 300  
DB 403 RDFSMSPIPSQFELPREYNRFHMHLEQSWRAYRDKLFWTHCVLATIMFTIFSFF 462  
QY 301 AEQLIALKRKIPPRRIHKEASPNRIRV 328

Db 463 AEQIALKRKIFPRRIHKEASPNRIRV 490

RESULT 2

Q9ULL2 PRELIMINARY; PRT; 950 AA.

AC Q9ULL2

DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)

DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)

DE KIAA1208 protein (Fragment).

GN Name=KIAA1208;

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Homidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

EX MEDLINE=20039619; PubMed=10574462;

RA Nagase T., Ishikawa K., Kikuno R., Hirose M., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. XV.

RT The complete sequences of 100 new cDNA clones from brain which code

RT for large proteins in vitro."

RL DNA Res. 6:337-345(1999).

DR ENBL; AB033034; BAR86522.2; -

DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0005509; P:calcium ion binding; IEA.

DR GO; GO:0030154; P:cell differentiation; IEA.

DR InterPro; IPR010506; DMAP binding.

DR InterPro; IPR02048; EF-hand.

DR InterPro; IPR00800; Notch region.

DR Pfam; PF06464; DMAP\_binding; 1.

DR Pfam; PF00036; ehand; 1.

DR SMART; SM00004; NL; 2.

DR PROSITE; PS00018; EF\_HAND; UNKNOWN\_1.

FT NON TER 1

SQ SEQUENCE 950 AA; 108785 MW; F653BDF6AEAC7503 CRC64;

Query Match 100.0%; Score 1731; DB 2; Length 950;

Best Local Similarity 100.0%; Pred. No. 4.7e-120;

Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPMPHMDIRVIMQELQDMFPPEFDKTSFHKVRS 60

DB 623 DTFADSLRYVNVKILNSKFGFTSRKVPAPMPHMDIRVIMQELQDMFPPEFDKTSFHKVRS 682

QY 61 EDMQFAFSFYFYLMSAVQPLNISQVDFVDTQSGVLSDRIRTLATRIHELPLSLQDLT 120

DB 683 EDMQFAFSFYFYLMSAVQPLNISQVDFVDTQSGVLSDRIRTLATRIHELPLSLQDLT 742

QY 121 GLEHMLNCSKMLPADITQLNNIPPTQESYDNLPPVTKSLVNTCKPVTDKIHAYKDK 180

DB 743 GLEHMLNCSKMLPADITQLNNIPPTQESYDNLPPVTKSLVNTCKPVTDKIHAYKDK 802

QY 181 NKYRFEIMGEIEIAFKMIRTNVSHVVGQDDIRKNPKRFVCLNDNDHNHDKDAQTVKAVL 240

DB 803 NKYRFEIMGEIEIAFKMIRTNVSHVVGQDDIRKNPKRFVCLNDNDHNHDKDAQTVKAVL 862

QY 241 RDPFESMFPISQFELPREYRNRLHMHLEQWRAEDKLFTHCVLATLIMFTIFSFF 300

DB 863 RDPFESMFPISQFELPREYRNRLHMHLEQWRAEDKLFTHCVLATLIMFTIFSFF 922

QY 301 AEQIALKRKIFPRRIHKEASPNRIRV 328

DB 923 AEQIALKRKIFPRRIHKEASPNRIRV 950

RESULT 3

Q6P9R5 PRELIMINARY; PRT; 921 AA.

ID Q6P9R5

Q6P9R5

DT 05-JUL-2004 (TREMBlrel. 27, Created)

DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)

DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)

DE Hypothetical protein.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6; TISSUE=Brain;

EX MEDLINE=22386257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.D., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Lequellano N.A., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Heiton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bourfard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,

RA Krzywinski M.I., Skalska U., Smalish U., Schnerch A., Schein J.E.,

RA Jones S.J., Marra M.A.

RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences."

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6; TISSUE=Brain;

RA Strausberg R.;

RL Submitted (Oct-2003) to the ENBL/GenBank/DBJ databases.

DR ENBL; BC060638; AAH0638.1; -

DR InterPro; IPR002048; EF-hand.

DR InterPro; IPR010983; EF-hand like.

DR InterPro; IPR00800; Notch region.

DR Pfam; PF00036; ehand; 1.

DR Pfam; PF00066; Notch; 1.

DR SMART; SM00004; NL; 2.

DR PROSITE; PS00018; EF\_HAND; UNKNOWN\_1.

KW Hypothetical protein.

SQ SEQUENCE 921 AA; 104970 MW; 12879CDE978D87BB CRC64;

Query Match 98.1%; Score 1698; DB 2; Length 921;

Best Local Similarity 97.6%; Pred. No. 1.3e-117;

Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVNVKILNSKFGFTSRKVPAPMPHMDIRVIMQELQDMFPPEFDKTSFHKVRS 60

DB 594 DTFADSLRYVNVKILNSKFGFTSRKVPAPMPHMDIRVIMQELQDMFPPEFDKTSFHKVRS 653

QY 61 EDMQFAFSFYFYLMSAVQPLNISQVDFVDTQSGVLSDRIRTLATRIHELPLSLQDLT 120

DB 654 EDMQFAFSFYFYLMSAVQPLNISQVDFVDTQSGVLSDRIRTLATRIHELPLSLQDLT 713

QY 121 GLEHMLNCSKMLPADITQLNNIPPTQESYDNLPPVTKSLVNTCKPVTDKIHAYKDK 180

DB 714 GLEHMLNCSKMLPADITQLNNIPPTQESYDNLPPVTKSLVNTCKPVTDKIHAYKDK 773

QY 181 NKYRFEIMGEIEIAFKMIRTNVSHVVGQDDIRKNPKRFVCLNDNDHNHDKDAQTVKAVL 240

DB 774 NKYRFEIMGEIEIAFKMIRTNVSHVVGQDDIRKNPKRFVCLNDNDHNHDKDAQTVKAVL 833

QY 241 RDPFESMFPISQFELPREYRNRLHMHLEQWRAEDKLFTHCVLATLIMFTIFSFF 300

DB 834 RDPFESMFPISQFELPREYRNRLHMHLEQWRAEDKLFTHCVLATLIMFTIFSFF 893

```
QY 301 AEQIALKPKIPRRRIHKEASPNRIRV 328
Db 894 AEQIALKPKIPRRRIHKEASPNRIRV 921

RESULT 4
AAH60638
ID AAH60638 PRELIMINARY; PRT; 921 AA.
AC AAH60638;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22389257; PubMed=12477932;
RA Strausberg R.J., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.B., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC060638; AAH60638.1; -
KW Hypothetical protein.
SQ SEQUENCE 921 AA; 104970 MW; 12879CDE978D87BB CRC64;

Query Match 98.1%; Score 1698; DB 2; Length 921;
Best Local Similarity 97.6%; Pred. No. 1.3e-117;
Matches 320; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYNKLNSKGFSTRKVPAPHPMHIDRIYVQELQDMFPPEFDKTSFKVRRHS 60
Db 594 DTFADSLRYNKLNSKGFSTRKVPAPHPMHIDRIYVQELQDMFPPEFDKTSFKVRRHS 653

QY 61 EDMQFAFSFYLLMSAVOPLNISQVFEVDVTDQSGVLSDRREIRTLRIHELPLSLQDLT 120
Db 654 EDMQFAFSFYLLMSAVOPLNISQVFEVDVTDQSGVLSDRREIRTLRIHDLPLSLQDLT 713

QY 121 GLEHMLNCSKMLPADITQLNIIPTQESYDDNLPVTKSLVTCNKVPVTDKHKAYKDK 180
Db 714 GLEHMLNCSKMLPANITQLNIIPTQEAAYDDNLPVTKSLVTCNKVPVTDKHKAYKDK 773

QY 181 NKYRFEIMGEBEIAFKMIRTNVSHVVGQDDIRKNPKRFVCLNDINDHNKDAQTVKAVL 240
Db 774 NKYRFEIMGEBEIAFKMIRTNVSHVVGQDDIRKNPKRFVCLNDINDHNKDAQTVKAVL 833

QY 241 RDFVESMPPIPSQFELPREYRNRLHMHLEQEWAYRDKLKFTHCVLATLIMFTIFSFF 300
Db 834 RDFVESMPPIPSQFELPREYRNRLHMHLEQEWAYRDKLKFTHCVLATLIIIFTIFSFF 893

RESULT 5
Q9NFW9
ID Q9NFW9 PRELIMINARY; PRT; 248 AA.
AC Q9NFW9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein DKFZp762B226 (Fragment).
GN Name=DKFZp762B226;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RA Bloembergen H., Boecher M., Brandt P., Mewes H.W., Weil B., Wiemann S.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL359588; CAB94874.1; -
DR PIR; T50618; T50618.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR010983; EF-hand like.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 248 AA; 29336 MW; D4C4CA6C4505522 CRC64;

Query Match 74.5%; Score 1290; DB 2; Length 248;
Best Local Similarity 100.0%; Pred. No. 5.6e-88;
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 VFEDVTDQSGVLSDRREIRTLRIHELPLSLQDLTGLEHMLNCSKMLPADITQLNII 144
Db 5 VFEDVTDQSGVLSDRREIRTLRIHELPLSLQDLTGLEHMLNCSKMLPADITQLNII 64

QY 145 PTQESYDDNLPVTKSLVTCNKVPVTDKHKAYKDKNKPFEIMGEBEIAFKMIRTNVSH 204
Db 65 PTQESYDDNLPVTKSLVTCNKVPVTDKHKAYKDKNKPFEIMGEBEIAFKMIRTNVSH 124

QY 205 VVGQLDDIRKNPKRFVCLNDINDHNKDAQTVKAVLRDQFYESMFPPIPSQFELPREYRNRF 264
Db 125 VVGQLDDIRKNPKRFVCLNDINDHNKDAQTVKAVLRDQFYESMFPPIPSQFELPREYRNRF 184

QY 265 LHMHELEQEWAYRDKLKFTHCVLATLIMFTIFSFFASOLIALKPKIPRRRIHKEASPN 324
Db 185 LHMHELEQEWAYRDKLKFTHCVLATLIMFTIFSFFAEQIALKPKIPRRRIHKEASPN 244

QY 325 RIRV 328
Db 245 RIRV 248

RESULT 6
Q9NV34
ID Q9NV34 PRELIMINARY; PRT; 132 AA.
AC Q9NV34;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ10959.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
```

```

RX PubMed=14702039;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Oobayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Puruya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuma N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kinata M., Watanabe M., Hirao K., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hotta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshina A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,
RA "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs."
RL Nat. Genet. 36:40-45 (2004).
DR EMBL; AK001821; BAA91926.1; -.
SQ SEQUENCE 132 AA; 16083 MW; 92B5DEB4E062AD07 CRC64;

Query Match 40.7%; Score 704; DB 2; Length 132;
Best Local Similarity 100.0%; Pred. No. 9.9e-45;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 MIRTNVSHVVGQDDIRKPRKFKVCLNDNDHNDKDAQTVKAVLRDFYSEMPPIPSQFEL 256
DB 1 MIRTNVSHVVGQDDIRKPRKFKVCLNDNDHNDKDAQTVKAVLRDFYSEMPPIPSQFEL 60

QY 257 PREYRNFLEHLEQWRAVYRDKLKFVTHCVLATLIMFTIFSFASQIALKRRKIPRRR 316
DB 61 PREYRNFLEHLEQWRAVYRDKLKFVTHCVLATLIMFTIFSFASQIALKRRKIPRRR 120

QY 317 IHKEASPNRIRV 328
DB 121 IHKEASPNRIRV 132

RESULT 7
Q61340 Q61340 PRELIMINARY; PRT; 384 AA.
AC Q61340;
DT 01-NOV-1996 (TrEMBLrel. 01; Created)
DT 01-NOV-1996 (TrEMBLrel. 01; Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26; Last annotation update)
DE Basic domain/leucine zipper transcription factor (Fragment).
GN Name=MafB;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX Cordes S.P., Barsh G.S.;
RA "The mouse segmentation gene kr encodes a novel basic domain-leucine
RT zipper transcription factor."
RL Cell 79:1025-1034 (1994).
DR EMBL; L36434; AAA65688.1; -.

```

```

DR PIR; I49528; I49528.
DR MGD; MGI:104555; MafB.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0005667; C:transcription factor complex; IDA.
DR GO; GO:0003677; F:DNA binding; IDA.
DR GO; GO:0005515; F:protein binding; IDA.
DR GO; GO:0045944; P:positive regulation of transcription from P...; IDA.
DR GO; GO:0007585; P:respiratory gaseous exchange; IMP.
DR GO; GO:0007379; P:segment specification; IMP.
DR InterPro; IPR010506; DMAP_binding.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR010983; EF_Hand_Like.
DR Pfam; PF06464; DMAP_binding; 1.
DR Pfam; PF00036; ehand; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 384 AA; 43753 MW; 51F473C8807A7E55 CRC64;

Query Match 33.6%; Score 581; DB 2; Length 384;
Best Local Similarity 97.4%; Pred. No. 5.3e-35;
Matches 112; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DTFADSLRYVKNILNSKFGFTSRKVPAPMHMIDRIYVQELQDMFPEEFKTSFHKVRS 60
DB 267 DTFADSLRYVKNILNSKFGFTSRKVPAPMHMIDRIYVQELQDMFPEEFKTSFHKVRS 326

QY 61 EDMQFAFSYFYLLMSAVQPLNISQVFDEVDTDQSGVLSDEIRTLATRIHELPLS 115
DB 327 EDMQFAFSYFYLLMSAVQPLNISQVFDEVDTDQSGVLSDEIRTLATRIHDLPLT 381

RESULT 8
Q7Q098 Q7Q098 PRELIMINARY; PRT; 717 AA.
AC Q7Q098;
DT 01-MAR-2004 (TrEMBLrel. 26; Created)
DT 01-MAR-2004 (TrEMBLrel. 26; Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26; Last annotation update)
DE AGCP9986 (Fragment).
GN Name=agCG49111; ORFNames=ENSAANG00000017588;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoides; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAA801008996; EAA00384.1; -.
FT NON_TER 1
SQ SEQUENCE 717 AA; 82794 MW; 0BA7D2B957E48250 CRC64;

Query Match 31.9%; Score 553; DB 2; Length 717;
Best Local Similarity 39.6%; Pred. No. 1.4e-32;
Matches 112; Conservative 57; Mismatches 104; Indels 10; Gaps 5;

QY 1 DTFADSLRYVKNILNSKFGFTSRKVPAPMHMIDRIYVQELQDMFPEEFKTSFHKVRS 60
DB 399 DTFADSLRYVKNILNSKFGFTSRKVPAPMHMIDRIYVQELQDMFPEEFKTSFHKVRS 458

QY 61 EDMQFAFSYFYLLMSAVQPLNISQVFDEVDTDQSGVLSDEIRTLATRIHELPLS 120
DB 459 EDMQFAFSYFYLLMSAVQPLNISQVFDEVDTDQSGVLSDEIRTLATRIHDLPLT 518

QY 121 GLEHMLNCS---KMLPADITQ-LNNIPPT--QESYYDENLPPVTKSLVNTCKPVTDKI 173
DB 519 YFEWVNTCSLQNNFLPLEHQQEQRQVYPLVYVEREDSTIPVTRSLVVGCAELADM 578

```





```

ID Q8SX14 PRELIMINARY; PRT; 666 AA.
AC Q8SX14;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE R325033p.
GN ORFNames=CG8027;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Ceiniker S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY089618; AAL90356.1; -
DR FlyBase; FBgn0033392; CG8027.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR008000; Notch_region.
DR Pfam; PF00066; Notch; 1.
SQ SEQUENCE 666 AA; 77745 MW; F6FDB6D1C1C39248 CRC64;
Query Match 29.2%; Score 506; DB 2; Length 666;
Best Local Similarity 33.6%; Pred. No. 4e-29;
Matches 108; Conservative
Yy 1 DTFADSLRYNKILSKFGTGRKVPAPMHPMDRIVMQLQDMFEEFDKTSFKVHRIS 60
Db 348 DIYSHSLIATNMLNLRAYGFKARHLVAHVGLDIDKQIVAMQRHFQQLDIAHQFRAP 407
Yy 61 EDQAFASYFYVLMASVQPLNISQVFEVDITDQSVLSREITLATRIHELPLSLQDLT 120
Db 408 TDQYAFASYFLMSETKVMSVEIFEPTDQSATWSDEVTFTLRIYQPLDWSNR 467
Yy 121 GLEHMLNCSK---MLPADITQLNNIPPTQSYDNPFPVTKSLVTVNCKPVTDKIHA 176
Db 468 YFEVQNCNRLNGLMKVDITVEHSTL--VYEREDSNLPTITRDJVVRCPLIAEALAN 525
Yy 177 YKDKYKRFEMGE--EIAFKMRTNVSHVVGQLDDIRKPKFKVCLNDNTDHNH-KDA 233
Db 526 FAVRPKYNFVSPKTSNFMMLTSLNTEVSDRLRLRKNPKFNCINDNDANRGEN 585
Yy 234 QTVKAVLRDEYSEFPFIPPSQFELPREYRNFRLHMLQEWRAVYRDKLKFTHCVLATLM 293
Db 586 EMVEHLLDFYLSFFPRSRKFELPPQYRNRFESWRDQWRK-RKRAVLVIGVGSLLLV 644
Yy 294 FTIPSPFAEQIALKRIKIFR 314
Db 645 VCLLRFNCHKAKLVRECVOR 665

RESULT 11
RA50_YEAST
ID RA50_YEAST STANDARD; PRT; 1312 AA.
AC P12753;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE DNA repair protein RAD50 (153 kDa protein).
GN Name=RAD50; OrderedLocusNames=YNL250W; ORFNames=N0872;
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OX NCBI_TaxID=4932;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=RE821;
RX MEDLINE=89276917; PubMed=2659437;
RA Alani E., Subbiah S., Kleckner N.;
RT "The yeast RAD50 gene encodes a predicted 153-kD protein containing a
RT purine nucleotide-binding domain and two large heptad-repeat
RT regions.";
RL Genetics 122:47-57(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=97377992; PubMed=9234673;
RA Sen-Gupta M., Gueldeger U., Beinbauer J.D., Fiedler T.A.,
RA Hegemann J.H.;
RT "Sequence analysis of the 33 kb long region between ORC5 and SUI1 from
RT the left arm of chromosome XIV from Saccharomyces cerevisiae.";
RL Yeast 13:849-860(1997).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site.
CC -!- SUBUNIT: Forms a complex with MRE11.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X14814; CAA32919.1; -
DR EMBL; X96722; CAA65494.1; -
DR EMBL; Z71526; CAA96157.1; -
DR PIR; S05808; BWSYDL.
DR Germonline; 143256; -.
DR SGD; S0005194; RAD50.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003690; F:double-stranded DNA binding; IEA.
DR GO; GO:0005515; P:protein binding; IPI.
DR GO; GO:0006303; P:double-strand break repair via nonhomologous...; IMP.
DR GO; GO:0000722; P:telomerase-independent telomere maintenance; IMP.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR007517; Rad50_zn_hook.
DR Pfam; PF04423; Rad50_zn_hook; 1.
DR TIGRFAMs; TIGR00606; rad50; 1.
KW ATP-binding; Coiled coil; DNA repair; Hydrolase; Meiosis.
FT NP_BIND 34 41 ATP (Potential).
FT DOMAIN 185 347 Coiled coil (Potential).
FT DOMAIN 403 558 Coiled coil (Potential).
FT DOMAIN 617 672 Coiled coil (Potential).
FT DOMAIN 734 1108 Coiled coil (Potential).
SQ SEQUENCE 1312 AA; 152568 MW; 58A0AA173AC5677E CRC64;
Query Match 7.2%; Score 124; DB 1; Length 1312;
Best Local Similarity 21.3%; Pred. No. 2.5;
Matches 68; Conservative 65; Mismatches 136; Indels 60; Gaps 14;
Yy 5 DSLRYVNVKILSKFGTGRKVPAPMHPMDR-----IVMQELQ---DMFPEEFD--KTSF 54
Db 189 DNLKSIKQMSVDIKLAKQSV-EHLKLDKRSKAMKLNHQLQTKIDQYNEVSEISQL 247
Yy 55 HKVHSEDMQAFSY-FYVLMASVQPLNISQVFEVDITDQSVLSREITLATRIHEL 113
Db 248 NEITEKSDKLFKSQDFQKILSKVENLNTKL-----SISD-QVKRLNSIDILD 296
Yy 114 LSLQDLTGLEHMLNCSKMLPADITQLNNIPPTQSYD--PNLPVTKSLVTVNCKPVD 171
Db 297 LSKPD---LQNLNLFKVLMDKKNQRLDLETDISLKDQSSLSLSLIRRQGL-- 351
Yy 172 KIKHAYKDKNKKYRFEIMGEEIAPMTIRTNVSHV-----VGQLDIRKKNPKFVC--LNDN 225

```

```
Db 352 EAGKETEKRNHLSLKE---AFCHKFGQLSNISNDMAQVNHMSQFKAFISQDLTDT 408
Qy 226 IDHNHKAQTVKAVLRDFYSMPFIPSPQFELPRYRNRLH- 266
Db 409 IDQFAKDIQLKETNLSDIKSIITVDSQNLVKNKDRSKLIHDSBELAELKLSFKSLSTQD 468
Qy 267 --MEHQWRAYRDKLKF 283
Db 469 SLNHELENLTKYKELQSW 487

RESULT 12
Q6FKKO PRELIMINARY; PRT; 691 AA.
AC Q6FKKO;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Similar to tr|Q12033 Saccharomyces cerevisiae YOR275c.
GN ORFNames:CAGJ0L109569;
OS Candida glabrata (Yeast) (Torulopsis glabrata).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
CX NCBI_TaxID=5478;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS138;
RG GENOLEVURES;
RA DuJon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boissame A., Boyer E., Cattolico L., Confanioleri F., de Daruvar A.,
RA Despons L., Fabre E., Faithhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swennene D., Tekalia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
RD EMBL; CR380958; CAG62218.1; -.
SQ SEQUENCE 691 AA; 80473 MW; 12D26888BDB8ABBC CRC64;

Query Match 6.8%; Score 117.5; DB 2; Length 691;
Best Local Similarity 19.4%; Pred. No. 3.4;
Matches 61; Conservative 54; Mismatches 128; Indels 71; Gaps 13;

Qy 23 RKVPAH-----MPHMIDRVNMQELQDMPEEFDKTSFHKVHSDMDFASY-- 69
Db 305 KHVESHQFSDFKVLYKLDKMDLIVDPDTEIAPSMLDQLFNLLPIIMYGTSYNE 364
Qy 70 ---FYLLMSAVQPLNISQV-----FDEVDTQSGVLSDRIRTLATRIHELPLSLQD 118
Db 365 KQDIYVLQHFVEPINLNIQLDELAGSFFEID-PQHIWNAHNKTSFVTR-KELDVIGQS 422
Qy 119 LTGHEHLNCSKMLPADITQLNNIPTQBSYDNPPLPVTKS-----LVTNCKPVT 170
Db 423 FGDLESNAINTE-----TQLANI-----TSYLDLEIKTYEEDKQQGVGVWTVTDARDVT 471
Qy 171 DKIHKAYDKNKKY-----RFEIMGEEIAFKMIRTNVSHVGVQLDD-----I 212
Db 472 KEFYEKVEKLQYLEIGRNINLELQELFASIDKSLVTSRTSKAHLTAQYNDPFLNKIEAL 531
Qy 213 RKNPRKFCVCLNDNDHMHKDAQTVKAVLRDFYSMFPFIPSPQFELPRYRNRLHMEHQE 272
Db 532 KKRERFI---SEIEKSYENRIITSLKIM-YKETDDIPGTISEREIFDNIYSKH---- 583
Qy 273 WRAYRDKLKFWTHC 286
```

```
Db 584 MKVFAVDMKYIEDC 597

RESULT 13
Q6JUT9 PRELIMINARY; PRT; 384 AA.
AC Q6JUT9;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE PDD4p.
GN Names:PDD4;
OS Pichia angusta (Yeast) (Hansenula polymorpha).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
CX NCBI_TaxID=4905;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS4732;
RA Monastyrzka I., Kiel J.A.K.W., Veenhuis M.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY304543; AAQ74772.1; -.
SQ SEQUENCE 384 AA; 45251 MW; E294618BBCE6AC12 CRC64;

Query Match 6.6%; Score 113.5; DB 2; Length 384;
Best Local Similarity 20.5%; Pred. No. 3.2;
Matches 69; Conservative 62; Mismatches 130; Indels 75; Gaps 15;

Qy 5 DSLRYVYNKILSKFGFTGRKVPAPHPHMIDRVNMQEL--QDMFPEEFDKTSFHKVHSD 62
Db 58 NAAQVYNQKLSAK-GYFSNK-----PNKLRQLLLSLDSSQLIPENAEFNFEISEKYVND 111
Qy 63 MQFASYSFYLLMSAVQPLNISQVDE-----VDTQSGVLSDRIRTLATRIHELPLSL 116
Db 112 RN-VNNIYSLNSAE--ISKQFRETALKKVAADKAEIELLNKEVEKLNKKVDEKERNI 167
Qy 117 QDLTGLEHMLINC-SKMLPADITQLNNIPTQBSYDNPPLPVTKSLVNTCKPVTDKIHK 175
Db 168 QSLQ-LDKIKSLDQAKHYCAMLNVNNAKNGEQERTFLYAEVKEELARNEMEV-DRLES 225
Qy 176 AYKDNKVRFEIMGEEETAFK-----MIRTN-----VSHVGVQLDDIRKPRK 218
Db 226 RLSNNMKNRASPRDDAAEAPKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 285
Qy 219 FVCLNDN-----IDHNHKAQTVKAVLRDFYSM-----FPIPS 252
Db 286 FQGLNSNKKVPSTPIPTDSELLAIDHNFDFLOKYDLILFDLLKPLNGSIDDTFFKKNX 345
Qy 253 QPELPRE---YRNRFHMHX-----LQEWYR 277
Db 346 QTDKQEIATSLNRNLSEMEQNYERVLATWQWKTYS 381

RESULT 14
AAQ74772 PRELIMINARY; PRT; 384 AA.
AC AAQ74772;
DT 01-JUN-2004 (TrEMBLrel. 27, Created)
DT 01-JUN-2004 (TrEMBLrel. 27, Last sequence update)
DT 01-JUN-2004 (TrEMBLrel. 27, Last annotation update)
DE PDD4p.
GN PDD4;
OS Pichia angusta (Yeast) (Hansenula polymorpha).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
CX NCBI_TaxID=4905;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS4732;
RA Monastyrzka I., Kiel J.A.K.W., Veenhuis M.;
RL "The H. polymorpha PDD4 gene product is required for selective
RT peroxisome degradation.";

```

